NBI-HE-96-35
July 1996
Revised January 1997

THE METRIC ON THE SPACE OF YANG-MILLS CONFIGURATIONS

Peter Orland[1]

The Niels Bohr Institute
Blegdamsvej 17, DK-2100
Copenhagen Ø, Denmark
orland@alf.nbi.dk

and

The City University of New York
Baruch College
and
The Graduate School and University Center
New York, NY
orland@gursey.baruch.cuny.edu [2]

Abstract

A distance function on the set of physical equivalence classes of Yang-Mills configurations considered by Feynman and by Atiyah, Hitchin and Singer is studied for both the $2+1$ and $3+1$-dimensional Hamiltonians. This set equipped with this distance function is a metric space, and in fact a Riemannian manifold as Singer observed. Furthermore, this manifold is complete. Gauge configurations can be used to parametrize the manifold. The metric tensor without gauge fixing has zero eigenvalues, but is free of ambiguities on the entire manifold. In $2+1$ dimensions the problem of finding the distance from any configuration to a pure gauge configuration is an integrable system of two-dimensional differential equations.

A calculus of manifolds with singular metric tensors is developed and the Riemann curvature is calculated using this calculus. The Laplacian on Yang-Mills wave functionals has a slightly different form from that claimed earlier.

In $3+1$-dimensions there are field configurations an arbitrarily large distance from a pure gauge configuration with arbitrarily small potential energy. These configurations resemble long-wavelength gluons. Reasons why there nevertheless can be a mass gap in the quantum theory are proposed.

[1]Work supported by PSC-CUNY Research Award Program grants nos. 666422 and 667438.
[2]Permanent address

1 Introduction

The Schrödinger representation is certainly the most powerful for visualizing and solving finite-dimensional quantum systems. This is undoubtedly why it is emphasized more than the Heisenberg and interaction representations, even in advanced textbooks on quantum mechanics. It is is just as useful in field theory as the interaction representation or path integrals for developing the rules of perturbation theory (though somehow this fact is not as often emphasized in textbooks) [1]. There is also a conceptual value to the notion of a quantum wave functional [2]. For a long time there has been the hope that a variational method might give a reasonable approximation to the vacuum of the Yang-Mills Hamiltonian [3]. Even should this turn out not to be the case, wave functionals of field configurations are easier to visualize than the abstact states of the Heisenberg representation and are worth investigating further.

Serious attempts to understand the long-distance physics of Yang-Mills theories in the Schrödinger representation were made by Feynman (in $2 + 1$ dimensions) [4] and by Singer [5]. These papers concerned the geometrical properties of the space of the physical configurations of the gauge field. This infinite-dimensional space is different in character from that of simpler field theories. In particular, two gauge fields which differ by a gauge transformation (one has the *option* of insisting that this gauge transformation is of Chern-Simons number zero in $3+1$ dimensions) must be identified as the same point of configuration space.

Feynman discussed the notion of distance between two field configurations [4]. He attempted to minimize this distance by making gauge transformations. In this way, he obtained an estimate for the distance between a pure gauge configuration and a configuration with a constant magnetic field. The point of doing this was to estimate the electric energy of certain wave functionals and thereby to minimize the total energy. On the basis of this result, he gave a conjecture for the lightest glueball state of the $2 + 1$-dimensional theory. In fact, the minimal distance had been already discussed by Atiyah, Hitchin and Singer, who used it to calculate the dimension of the moduli space of instantons [6].

Singer defined an inner product on the configuration space and discussed how to regularize tensor contractions so that the eigenvalues of the Ricci tensor and the Laplacian (which is the kinetic term of the Yang-Mills Hamiltonian) were well-defined [5]. The first work along similar lines can be found in the papers by Babelon, Viallet, Mitter, Narasimhan and Ramadas [7, 8]. In particular, Babelon and Viallet studied geodesics on this space (and thereby classical dynamics with no potential energy) [8]. A difficulty encountered in [8] was that geodesic motion appeared ill-defined at points of the manifold where the operator $\mathcal{D}^2$ has zero eigenvalues.

Recently an important advance was made by Karabali and Nair [9], who discovered a gauge-invariant description of Yang-Mills theories in $2+1$ dimensions. Their method solves the constraint of Gauss' law and they have been able to show the existence of a gap at strong coupling. They have also made some arguments for the persistence of this gap at any coupling [10].

In this article, the configuration space of the Yang-Mills theory is investigated using general methods, some of which are not specific to the number of dimensions. The issue

of mass generation is confronted, but not that of color confinement.

After introducing the notation and reviewing the relevant physics in section 2, a heuristic way to eliminate Gauss' law from the $A_0 = 0$ Yang-Mills theory is discussed in section 3. The methods used to study gauge theories in references [5, 7, 8] start from the same standpoint. The difficulty raised in [8] concerning zero modes of the covariant operator $\mathcal{D}^2$ is shown. Because of this problem, such a formalism cannot be accepted on faith, and it is carefully justified in sections 4, 5 and 6. In the correct metric tensor, these zero modes do not present a problem.

A distance function motivated by Feynman's analysis [4] is investigated and found to define a metric space in section 4. This distance function coincides with that of Atiyah et. al. [6], though it will not be considered only on self-dual connections. Furthermore, with this metric, the space of physical configurations is complete, meaning roughly that there are no "holes" in this space. In section 5, it is found that this distance function is also the "best" choice, in that the distance between two points is the length of a minimal curve connecting these points. In fact, this property had been conjectured by Babelon and Viallet [8]. In section 6, the distance function is used to define the Riemannian manifold where the gauge connection itself is used as the coordinate. This is a very different choice of coordinates from that taken in references [9, 10, 11]. It is shown that this can be done everywhere on the manifold, in spite of the fact that the metric tensor has zero eigenvalues. The problem with zero modes of $\mathcal{D}^2$ [8] is completely resolved. Since singular metric tensors are rarely used, a review and a simple finite-dimensional example are given in the appendix. Gauss' law can be solved in any dimension of space-time. The Hamiltonian is not local (this is also the case for a physical gauge condition, such as the axial gauge, and also in a polar representation of the gauge field [12]). For non-infinitesmal distances, it is hard to calculate the distance function explicitly, but in section 7 it is shown that for some interesting cases in $2 + 1$ dimensions its calculation can be reduced to the solution of an integrable set of differential equations.

There are various theorems relating the eigenvalues of Ricci tensor to the spectrum of the Laplacian [13]. Ultimately, these results may be of value to the Yang-Mills Hamiltonian. The validity of these theorems usually depends upon the manifold having compact closure. This is probably not true of the unregularized Yang-Mills configuration space, but is certainly so if the theory is put on a finite lattice [14]. In section 8 the curvature tensor is found at points where zero modes are not present [5, 8], but no application of these results is made to the spectrum.

Once the metric on the manifold of Yang-Mills configurations is known, the kinetic-energy operator (in other words the electric energy) can be obtained. This operator, the coordinate-invariant Laplacian, is found in section 9. The result differs in two respects from that obtained by previous authors [5, 8]. First of all, the zero modes of $\mathcal{D}^2$ present a complication, which leads to a slightly different metric tensor. Secondly, even with the correct metric tensor, an additional term is present in the Laplacian, which had been missed in references [5, 8].

In section 10 some unusual aspects of the magnetic energy functional are investigated for different dimensions of space-time. In $3 + 1$ dimensions, for at least some regions of the manifold of configurations, the "hills and valleys" of the potential energy

are extremely steep. In particular, there are regions a small geodesic distance from any pure gauge of extremely large potential energy. What is even more striking is that there exist points on the manifold of arbitrarily large geodesic distance from any pure gauge, yet possess arbitrarily small potential energy; this is very surprizing for a theory which is expected to have a mass gap in the spectrum. In contrast, the potential energy of the $2+1$-dimensional model can have *any* positive value for configurations a fixed distance from the pure gauge configuration. Discussion of the origin of the mass gap for $2+1$ and $3+1$ dimensions (in spite of the difficulty noted above!) is in Section 11.

The philosophy here differs significantly from that of most people currently working in this field [15, 16, 17, 18, 19] in one significant respect. The space of Yang-Mills connections (discussed here for only connections in space with $A_0 = 0$, rather than for the Euclidean path integral) is not *restricted* to configuration space (or orbit space) by some sort of gauge fixing. Instead the points of configurations space are defined as equivalence classes of connections, and no restrictions whatever are imposed on these connections. In this sense, the approach here is similar in spirit to that of Friedberg et. al. [20]. The problem of finding the fundamental region (or the interior of the Gribov horizon) has been traded for a singular metric tensor. Metric tensors with vanishing eigenvalues introduce subtle new features into differential geometry, but as shown in the appendix, they present no inconsistencies. So-called reducible connections [19], i.e. gauge configurations which are invariant under a subgroup of gauge transformations present no difficulty in defining the Laplacian on configuration space. It is argued that they are also not relevant in the calculation of inner products between certain wave functionals (though this issue needs to be examined more carefully). This will be briefly discussed at the end of section 6.

Since the literature on the configuration space of Yang-Mills gauge theories is by now quite extensive, it is important to say at the outset what in this paper can or cannot be found elsewhere. The ideas discussed in section 3 are not new and can be found in references [5, 7, 8]. While many of the ideas of sections 4, 5 and 6 can be found in references [4, 5, 6, 7, 8], the ideas are tied together in a way which is entirely new and the main results are slightly different. Most importantly, it is shown that the metric tensor and geodesic trajectories are well-defined *even at so-called non-generic points* (where there are zero modes of $\mathcal{D}^2$). Singer presented the curvature tensor for other points in his paper [5], as did Babelon and Viallet [8]. This tensor is calculated using the methods introduced in this paper in section 8, though the result of this calculation is not new. The kinetic term of the Yang-Mills theory obtained in section 9 differs in certain respects from that claimed earlier [5, 8]. As far as I know, none of the results presented in sections 7, 9, 10 and 11 can be found elsewhere.

As I have found the properties of the manifold of physical Yang-Mills configurations hard to visualize, I have tried to use careful definitions and theorems in sections 4, 5 and most of section6, so as not to prove wrong results. These sections use some basic concepts of real analysis and integration theory, but are hopefully accesible to those physicists who are not well-versed in these concepts. No attempt to be completely rigorous is made in rest of the paper.

2 Preliminaries

The main purpose of this section is to establish some notation, conventions and definitions which will be used in the remainder of the article. Those initiated into the Hamiltonian formulation of gauge theories will not find anything particularly new until section 3.

The dimension of space-time is denoted by $D+1$. Let t_a be an orthonormal basis for the fundamental repesentation $su(N)$, the Lie algebra of the group $SU(N)$. This basis can be chosen so that $tr\, t_a t_b = \delta_{a\,b}$ and $[t_a, t_b] = f^{a\,b\,c}\, t_c$. A gauge field A is an $su(N)^D$-valued vector field $A_j(x)$, $j = 1, ..., D$ on a D-dimensional manifold M, whose points will be denoted by letters x, y, z. The covariant derivative operator for the gauge field A is $D_j^A = \partial_j - iA_j$, which will sometimes just denoted by D_j, depending on the context. The adjoint representation of the covariant derivative is $\mathcal{D}_j \equiv [D_j, \cdot]$. The curvature or field strength is $F_{jk}^A = -i[D_j^A, D_k^A]$.

A gauge transformation is any differentiable mapping of M into the fundamental representation of $SU(N)$. The space of gauge transformations will be written G (in other words $h \,\epsilon\, G$ means $h(x) \in SU(N)$). Let $A_\mu(x) = A_\mu^a(x)\, t_a \in su(N)$ be a gauge field. If $B_\mu(x) = B_\mu^a(x)\, t_a \in su(N)$ such that A can be changed to B by a gauge transformation $h \in G$, i.e. $h(x)^{-1}\, D_\mu^A(x)\, h(x) = D^B$, then A and B will be said to be gauge equivalent, sometimes expressed as $B = A^h$ (a more precise definition of gauge equivalence will be given in section 4). Clearly gauge equivalence is an equivalence relation. For $D \neq 3$ two gauge-equivalent fields will also said to be physically eqivalent.

For the case of $D = 3$ a subtlety arises which permits (but does not require) the definition of a different equivalence relation [21]. With this definition, a pair of gauge-equivalent gauge fields are not neccesarily physically equivalent. Suppose that M is taken to be R^3 with all gauge fields falling off at infinity as

$$A_j(x) \longrightarrow -i\; h(x)^{-1} \partial_j h(x) + O(1/x)\;.$$

A gauge transformation $h \,\epsilon\, G$ has an associated Chern-Simons number

$$\nu[h] = \frac{i}{12\pi^2} \int d^3x\; \varepsilon^{i\,j\,k}\; tr\; h(x)^{-1} \partial_i h(x)\; h(x)^{-1} \partial_j h(x)\; h(x)^{-1} \partial_k h(x)\;, \qquad (1)$$

which is an integer. One has the option of insisting that two gauge configurations A and B are physically equivalent only if $h(x)^{-1}\, D_\mu^A(x)\, h(x) = D^B$, with $\nu[h] = 0$. Now if $h, f \,\epsilon\, G$ are two gauge transformations, then $\nu[hf] = \nu[h] + \nu[f]$. This property means that :

- $\nu[h] = \nu[f]$ is an equivalence relation and therefore the gauge transformations with a particular Chern-Simons number ν constitute an equivalence class G_ν and
- physical equivalence of gauge fields is an equivalence relation.

Gauge configurations have the Chern-Simons integral

$$C[A] = \frac{1}{12\pi^2} \int d^3x\; \varepsilon^{i\,j\,k} tr(A_i F_{j\,k} + \frac{2i}{3} A_i A_j A_k)\;, \qquad (2)$$

which is changed under a gauge transformation satisfying (1) by $C[A] \to C[A^h] = C[A]+\nu[h]$. Thus the Chern-Simons integral is the same for each element of a physical-equivalence class making it sensible to write $C[A] = C[\alpha]$.

The Hamiltonian of the $SU(N)$ Yang-Mills theory in $A_0 = 0$ (temporal) gauge in the Schrödinger picture is

$$H = \int_M d^D x \, [-\frac{e^2}{2} tr \, \frac{\delta^2}{\delta \, A_j(x)^2} + \frac{1}{4e^2} tr \, F_{j\,k}(x)^2] \; . \tag{3}$$

The allowed wave functionals Ψ satisfy the condition that if A and B are physically equivalent

$$\Psi[A] = \Psi[B] \; . \tag{4}$$

The space of gauge field configurations (or connections) will be denoted by $\mathcal{U}$. The physical configuration space (sometimes called the orbit space) is then $\mathcal{U}/G \equiv \mathcal{M}_D$ when $D \neq 3$ and $\mathcal{U}/G_0 \equiv \mathcal{M}_3$ when $D = 3$. In other words, for $D \neq 3$, configuration space, $\mathcal{M}_D$ is defined to be the set of gauge-equivalence classes of gauge fields (these were already defined to be physically equivalent). For $D = 3$, configuration space $\mathcal{M}_3$ is defined to be the set of physical-equivalence classes of gauge fields. The elements of $\mathcal{M}_D$ will be called physical configurations. The physical configurations containing the gauge fields $A, B, S, ... \in \mathcal{U}$ will be respectively denoted $\alpha, \beta, \sigma, ... \in \mathcal{M}_D$. Thus, A is physically equivalent to B if and only if $\alpha = \beta$. By (4), wave functionals depend only on physical configurations (α in $\mathcal{M}_D$), not gauge configurations (A in $\mathcal{U}$).

3 Projecting out gauge transformations

Can one modify the form of the Hamiltonian (3) so that imposing (4) is unneccesary? A simple guess is to change the kinetic term to obtain

$$\begin{aligned} H = & -\int_M d^D x \, \frac{e^2}{2} tr \, \frac{\delta}{\delta \, A_j(x)} \; G^{-1}_{j\,k} \; \frac{\delta}{\delta \, A_k(x)} \\ & + \int_M d^D x \, \frac{1}{4e^2} tr \, F^2_{j\,k}(x) \; , \end{aligned} \tag{5}$$

where

$$G^{-1}_{j\,k} = \delta_{j\;k} - \mathcal{D}_j \frac{1}{\mathcal{D}^2} \mathcal{D}_k \; , \tag{6}$$

is a projection operator (being very sloppy about domains of covariant derivative operators) which vanishes on infinitesmal gauge transformations $h(x) \approx 1 - i\omega$, that is $G^{-1}_{j\,k} \mathcal{D}_k \omega = 0$. For $D = 3$, G^{-1} neccesarily vanishes on gauge transformations whose Chern-Simons number is zero.

The reader may have noticed a serious problem. The operator $\mathcal{D}^2$ can have eigenvectors with zero eigenvalues [8]. How should the Green's function $1/\mathcal{D}^2$ in (6) be interpreted in this case?

It is useful to introduce a generalization of the inverse of an operator. If Θ is any operator on a space with orthonormal basis $\{|X)\}$ and $\Theta|X) = \lambda_X|X)$, then the

operator and its generalized inverse have the spectral representations (again being very careless about domains)

$$\Theta = \sum_X \lambda_X |X)(X| \;,$$
$$\Theta^{-1} = \sum_{X, \lambda_X \neq 0} \lambda_X^{-1} |X)(X| \;. \tag{7}$$

In the event that $\mathcal{D}^2$ has zero modes [8], its inverse can be made well-defined using (7). However, it is not obvious that this is the correct prescription. It will be shown in section 6 that precisely (7) must be used to define the inverse of $\mathcal{D}^2$. This means that (6) is incorrect and must be replaced by

$$G^{-1}_{j\,k} = \delta_{j\,k} - \mathcal{D}_j \, \mathcal{P} \frac{1}{\mathcal{D}^2} \, \mathcal{D}_k \;, \tag{8}$$

where $\mathcal{P}$ denotes the principal value. This is equivalent to using (7) to define the inverse of $\mathcal{D}^2$. This may also be written as

$$\mathcal{P} \frac{1}{\mathcal{D}^2} = \frac{1}{2} \left(\frac{1}{\mathcal{D}^2 + i\epsilon} + \frac{1}{\mathcal{D}^2 - i\epsilon} \right) .$$

The operator G^{-1} defined in (6) or (8) is formally idempotent (it is equal to its square), hence, if it is meaningful, each of its eigenvalues can only be either zero or one. It does not have an inverse except in the sense of (7). According to this definition G^{-1} is its own inverse, $G = G^{-1}$; it plays the role of the identity on the space of connections modulo small gauge transformations.

A conjecture for the kinetic term is then:

$$T = -\frac{e^2}{2} \left(\frac{\delta}{\delta A} \right)^T G^{-1} \frac{\delta}{\delta A} \;, \tag{9}$$

This resembles a covariant Laplacian on a Riemannian metric space with metric tensor G, since the determinant of the idempotent operator G, upon removal of zero modes, is equal to one and $G = G^{-1}$. It turns out, however, that (9) is still the wrong answer, even with the correct metric (8). There is another term present.

Before taking these ideas seriously, several problems need to be dealt with. The problem of inverting $\mathcal{D}^2$ [8] must be solved. It needs to be verified that the complicated operator (8) has no zero modes other than small gauge transformations; other differential operators certainly do [22]. These problems will be solved in section 6. The Laplacian needs to be determined, as will be done in section 9. It is perhaps noteworthy that the Hamiltonian of the form (5) is easily shown to be correct for an Abelian gauge theory.

The reader may wonder why is is really valuable to study the Laplacian in the approach of this paper and references [4, 5, 7, 8], when a perfectly good definition of the quantized Yang-Mills theory already exists; the lattice gauge theory. On the lattice, at strong coupling, it is possible to implement Gauss' law to obtain only physical states, at least on a case by case basis; they are strings. However, even with a lattice theory,

it is difficult to get much intuition about the weak coupling spectrum. An axial gauge eliminates the Gauss' law constraint, but at the cost of introducing difficult boundary conditions and an unmanageable kinetic term. The methods introduced here are not intended to compete with the lattice, but rather to have a better understanding of the geometry of the manifold $\mathcal{M}_D$ and the potential energy function on this manifold. A lattice discussion should be very useful and will appear in reference [14].

In the next section a distance function will be defined; in section 6 it will be shown that the infinitesmal form of this distance function exists and that the metric tensor is G, as defined in (8). There are various theorems relating the eigenvalues of Ricci tensor to the spectrum of the Laplacian [13]. Ultimately, these results may be of value to the Yang-Mills Hamiltonian. The validity of these theorems usually depends upon the manifold having compact closure. This is probably not true of the unregularized Yang-Mills configuration space, but is certainly so if the theory is put on a finite lattice [14]. The curvature is is found at points where zero modes are not present [5, 8] in section 8. The Laplacian is determined in section 9.

4 The metric space of configurations

This section and sections 5 and 6 presuppose some knowledge of unsophisticated aspects of analysis on the part of the reader, such as completeness, total boundedness, compactness, measure and Hilbert spaces [23]. However, it is hoped that most theoretical physicists, who usually do not feel the need of such concepts (and who perhaps have happily forgotten them) will be able to follow the line of argument. The rigorous part of this article begins here (it ends in section 6).

First a distance function will be defined [4, 6]. Feynman found an estimate of this distance function between the pure gauge configuration and a configuration of constant magnetic curvature for $D = 2$. The discussion here will be more general, and it will be shown that $\mathcal{M}_D$ equipped with this function is a metric space. This fact has already been noted for the case of self-dual connections [6]. The value of establishing this is that the distance function must be therefore a continuous function of the physical configurations. This is one of the essential ingredients needed to show that the manifold can be equipped with a Riemannian metric. Furthermore, this metric space is complete. It will be shown in section 6 that the metric tensor defined with this distance function is (8). A general reference on metric spaces from the geometric point of view is the book by Aleksandrov and Zalgaller [24].

Some more careful definitions are needed. The space of connections $\mathcal{U}$ is redefined to contain only those gauge fields which are Lebesgue measurable, and are square-integrable, i.e.

$$\int_M d^D x \sum_{k=1}^{D} tr\ A_k(x)^2 < \infty \ ,$$

Physically, one would like to also restrict the connections to those whose field

strengths exist and are square-integrable, i.e.

$$\int_M d^D x \sum_{j,k=1}^{D} tr\ F_{j\,k}(x)^2 < \infty\ .$$

However, making this further restriction will ruin the completeness property of the space of connections and is quite incovenient, as discussed in the next paragraph.

Two gauge fields A and $B \in \mathcal{U}$ are identified if they are the same except on a set of measure zero. In other words, no distinction is made between gauge fields which are the same almost everywhere. Thus the space $\mathcal{U}$ is a Hilbert space, and gauge fields are representative of vectors in this Hilbert space. Furthermore, G and G_ν do not consist of any $SU(N)$ valued functions $g(x)$, but only those which are differentiable and for which $ig^{-1}\partial g \in \mathcal{U}$. This means in particular that

$$\int_M d^D x \sum_{k=1}^{D} tr\ [ig(x)^{-1}\partial_k g(x)]^2 < \infty. \tag{10}$$

Any element of $\mathcal{U}$ is mapped into another element of $\mathcal{U}$ by such a gauge transformation. Notice that two almost-everywhere-equal gauge fields are still equal almost everywhere after such a gauge transformation is appiled to both. Restricting the integral of the magnetic field strength squared destroys the completeness property, and the space of connections is no longer a Hilbert space. However, the space of such restricted connections is dense in $\mathcal{U}$. Therefore it not neccesary to make this restriction. The situation is similar to that one encounters when studying unbounded operators (the subspace of a Hilbert space which is acted upon by an unbounded operator is dense in the Hilbert space [23]).

It is not quite enough to say that two vectors in $\mathcal{U}$ are gauge-equivalent if one can be gauge transformed to another by some gauge transformation satisfying (10). The equivalence classes must actually be made larger in order to obtain a metric space. Instead two vectors in $\mathcal{U}$ with representatives A and B will be said to be gauge-equivalent if there is a sequence of gauge transformations g_1, g_2,... each satisfying (10), such that

$$B = \lim A^{g_n}\ , \tag{11}$$

in the usual metric of the Hilbert space U, i.e. the square root of the integral of the square of the difference of A, $B \in \mathcal{U}$:

$$\|A - B\| = \sqrt{\frac{1}{2}\int_M d^D x\ tr\ [A^h(x) - B^f(x)]^2}$$

(the extra $\frac{1}{2}$ is included purely for convenience in the later discussion, and is irrelevant to the arguments in this paragraph). If the dimension of space is not equal to three, A and B are also defined to be physically equivalent. In three space and one time dimension, two connections A and B are said to be physically equivalent if (11) holds and the Chern-Simons index of each of the gauge transformations g_1, g_2,... is equal to zero.

Both of these new versions of gauge-equivalence and physical-equivalence are, in fact, equivalence relations. It is obvious that every connection is equivalent to itself. If (11) holds, then $A = \lim B^{g_n^{-1}}$. This is because for any $\epsilon > 0$, there exists a positive integer J, such that if $n > J$, $\epsilon > \|A^{g_n} - B\|$. But $\|B^{g_n^{-1}} - A\| = \|A^{g_n} - B\|$. The relations of gauge-equivalence and physical-equivalence are symmetric. To prove transitivity of gauge-equivalence, suppose that (11) is valid, and for some connection C,

$$C = \lim B^{h_n} ,$$

for some sequence of gauge transformations h_1, h_2,... (which must, by definition, satisfy (10)). Then, defining for each $n = 1, 2,...$ $f_n(x) = g_n(x)h_n(x)$, one can easily see that each f_n is a gauge transformation (i.e. satisfying (10)). The sequence A^{f_n} converges to C. To see this, note that for any $\epsilon > 0$ there exist positive integers J and K such that for $n > J$, $\|A^{g_n} - B\| < \epsilon/2$ and for $n > K$, $\|B^{h_n} - C\| < \epsilon/2$. So if n is larger than both J and K,

$$\|A^{f_n} - C\| \leq \|A^{f_n} - B^{h_n}\| + \|B^{h_n} - C\| = \|A^{g_n} - B\| + \|B^{h_n} - C\| < \epsilon .$$

It may seem odd to the reader that two gauge-equivalent connections are not neccesarily transformable to each other by gauge transformations. The reason for this is the differentiability requirement and the condition (10). In fact, if two connections are gauge equivalent and only one of them is continuous, it is clear that no "nice" (that is differentiable and square-integrable) gauge transformation can relate the two.

Let α and β be two physical configurations. Let $G_D = G$ for $D \neq 3$ and $G_D = G_0$ for $D = 3$. The distance function on $\mathcal{M}_D$ is defined by

$$\rho[\alpha, \beta] = \inf\{\sqrt{I[A, B; h, f]} \ : \ A \in \alpha, B \in \beta, h \in G_D, f \in G_D\} , \tag{12}$$

where

$$I[A, B; h, f] = \frac{1}{2} \int_M d^D x \; tr \; [A^h(x) - B^f(x)]^2 \tag{13}$$

$$= \frac{1}{2} \int_M d^D x \sum_{k=1}^{D} tr[ih(x)^{-1}\partial_k h(x) + h(x)^{-1} A_k(x) h(x)$$

$$- \ if(x)^{-1}\partial_k f(x) - f(x)^{-1} B_k(x) f(x)]^2 .$$

For the non-practitioner of real analysis, who may be discouraged by the notation, the right-hand-side of equation (12) means "the greatest lower bound of the set of values taken by $\sqrt{I[A, B; h, f]}$ for any $A \in \alpha$, $B \in \beta$, $h \in G_D$, and $f \in G_D$". Since the gauge connections are square integrable, the integral (13) always converges, but is not bounded from above. The distance $\rho[\cdot, \cdot]$ can be either zero or any positive real number.

Since for any $j \ \epsilon \ G$

$$I[A, B; h, f] = I[A, B; hj, fj] , \tag{14}$$

It is possible to simplify (12) to

$$\rho[\alpha,\beta] = \inf\{\sqrt{I[A,B;h]} \;:\; A \in \alpha,\ B \in \beta,\ h \in G_D\}\ , \tag{15}$$

where $I[A,B;h] \equiv I[A,B;h,1]$. However, sometimes (12) is more convenient to use than (15), as will be seen below.

The function $\rho[\cdot,\cdot]$ is called a metric on $\mathcal{M}_D$, if for any α, β and σ in $\mathcal{M}_D$

$$\rho[\alpha,\beta] \geq 0\ , \tag{16}$$

$$\rho[\alpha,\beta] = \rho[\beta,\alpha]\ , \tag{17}$$

$$\rho[\alpha,\beta] + \rho[\beta,\sigma] \ \geq\ \rho[\alpha,\sigma]\ , \tag{18}$$

$$\rho[\alpha,\beta] = 0 \quad \Longleftrightarrow \quad \alpha = \beta\ . \tag{19}$$

If these properties are satified, then $\mathcal{M}_{\mathcal{D}}$ together with $\rho[\cdot,\cdot]$ is called a metric space. By this definition some physical examples of manifolds for which "metrics" are defined are not really metric spaces (Minkowski space for example). Of these four properties, the two which are most obviously true are (16) and (17). Property (19) follows the definition of a Hilbert space vector as an equivalence class of almost-everywhere-equal functions (from the beginning "equal" gauge fields were defined to be those which are equal almost everwhere) and from the definition of physical equivalence using equation (11) (in fact the reason why the equivalence classes were enlarged was precisely so that (19) would be satisfied). In order to be able to prove some very basic analytic properties, e.g., that any convergent sequence must be a Cauchy sequence, and $\rho[\cdot,\cdot]$ a continuous function of its arguments, that the space is Hausdorff, etc. it is neccesary that the triangle inequality (18) be satisfied.

The validity of (18) is not entirely trivial. In general, coset spaces in Euclidean spaces do not satisfy the triangle inequality. As an example, consider a family of curves in R^2 (see figure 1). If two points $x, y \in R^2$ lie in the same curve, write $x =_C y$. Clearly $=_C$ is an equivalence relation. A distance between two equivalence classes (that is, two curves) can be defined as the minimum of the distance between a point on one curve and a point on the other. This distance does *not* satify the triangle inequality for an arbitrary choice of three elements of the coset space.

Proposition 4.1 : *The function $\rho[\cdot,\cdot]$ satifies the triangle inequality (18).*

Proof: Notice that for any gauge fields $A \in \alpha$, $B \in \beta$ and $S \in \sigma$ and any gauge transformations $h, f, j \,\epsilon\, G_D$

$$\sqrt{I[A,B;h,f]} + \sqrt{I[B,S;f,j]} \geq \sqrt{I[A,S;h,j]}\ .$$

Given any real $\epsilon > 0$, it is possible to chose h and f so that $\sqrt{I[A,B;h,f]} = \rho[\alpha,\beta] + \epsilon/2$ (by the definition of the greatest lower bound). By property (14) there exists a j such that $\sqrt{I[B,S;f,j]} = \rho[\beta,\sigma] + \epsilon/2$. Therefore

$$\rho[\alpha,\beta] + \rho[\beta,\sigma] \geq \sqrt{I[A,S;h,j]} - \epsilon \geq \rho[\alpha,\sigma] - \epsilon\ .$$

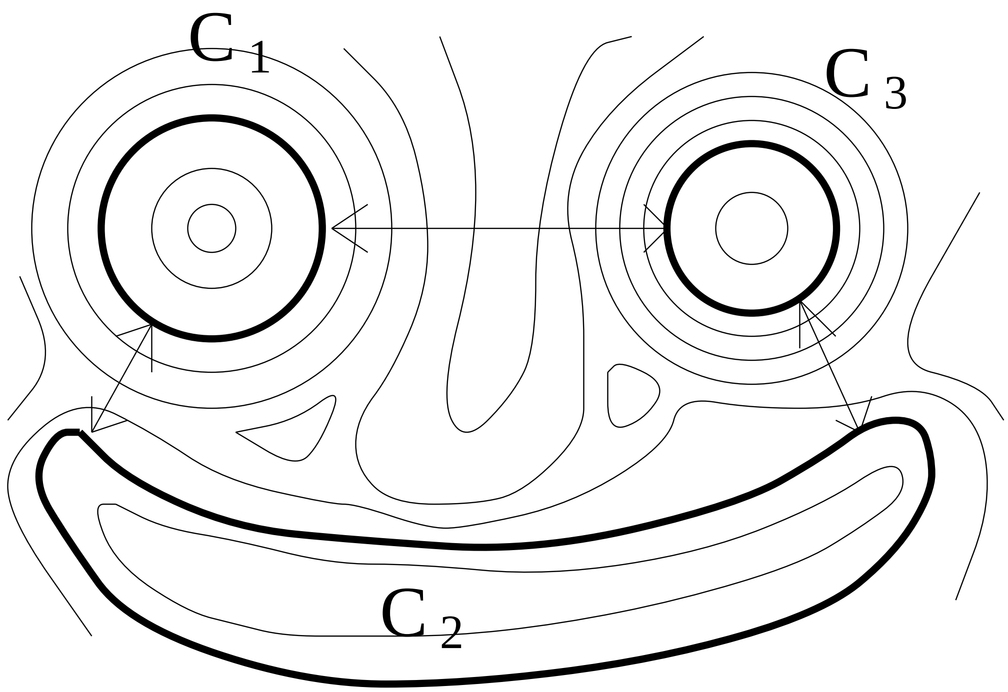

Figure 1: Consider a family of curves in the two-dimensional plane. If the distance between two curves is defined as the length of the shortest line connecting them, then this distance does not always satisfy the triangle inequality. Here the distance from curve C_1 to curve C_2 plus the distance from curve C_2 to C_3 is *less* than the distance from curve C_1 to C_3.

Since ϵ is arbitrary (18) follows.

A complete metric space is one for which any Cauchy sequence, i.e. a sequence where progressive points get closer to each other, always has some unique limit point [23] (the converse is always true by the triangle inequality). Thus a complete metric space has no pathological properties, such as isolated points or closed domains being "missing" from the space. More precisely, a sequence α_1, α_2,...,α_n,... is a Cauchy sequence if for any real $\epsilon > 0$, there exists an integer R such that if $n, m > R$, then $\rho[\alpha_n, \alpha_m] < \epsilon$.

Proposition 4.2 *: The metric space $\mathcal{M}_D$ is complete.*

Proof: Recall the metric on the space of gauge connections, $\mathcal{U}$ given above. For A, $B \in \mathcal{U}$, this is

$$\|A - B\| = \sqrt{I[A, B; 1, 1]}\,.$$

With this metric, $\mathcal{U}$ is a complete metric space (in fact, a Hilbert space, each whose vectors is representated by a set of measurable, square-integrable, Lie-algebra-valued functions, all equal except on a set of measure zero). Let α_1, α_2,...,α_n,... be a Cauchy sequence. Then for any $\epsilon > 0$ there exists some R such that for $n, m > R$, $\rho[\alpha_n, \alpha_m] < \epsilon/2$. From the definition of the metric and from property (14) it is possible to find

$A_1 \in \alpha_1$, $A_2 \in \alpha_2$,..., $A_k \in \alpha_k$, such that for $n, m > R$ $\|An - A_m\| = \rho[\alpha_n, a_m] + \epsilon/2 < \epsilon$. Thus, since $\|\cdot - \cdot\|$ is a metric on a complete metric space $\mathcal{U}$, the sequence A_1, A_2,...,A_n,... converges in this metric. Since $\rho[\alpha, \beta]$ is bounded above by $\|A - B\|$, the sequence α_1, α_2,...,α_n,... therefore converges in $\mathcal{M}_D$.

Though complete, the metric space $\mathcal{M}_D$ is not compact, for it is not totally bounded or even bounded. It is possible to make the space bounded in a variety of ways. For example, one can define

$$\rho_\Lambda[\alpha, \beta] = \inf\{\sqrt{I_\Lambda[A, B; h, f]} \ : \ A \in \alpha, B \in \beta, h \in G_D, f \in G_D\} \ ,$$

where

$$I_\Lambda[A, B; h, f] = \frac{1}{2} \int_M d^D x \ \frac{tr[A^h - B^f]^2}{\{1 + \frac{1}{\Lambda}\sqrt{tr[A^h - B^f]^2} \ \}^2} \ ,$$

It can be shown with some work that this distance function is a metric and that the metric space so defined is complete. The reader will notice that there is an upper bound on this distance: $\rho_{V,\Lambda}[\alpha, \beta] \leq V\Lambda^2/2$. However, while this metric space is bounded, it is not obviously totally bounded. I do not have a proof that the space is not totally bounded, but this possibility seems rather remote. This is unfortunate, for a space of compact closure is needed to prove many of the standard results on the spectra of coordinate-invariant Laplace operators [13] (I don't know which of these results are generally true without assuming compact closure). In fact, the only way I know of to make the space of gauge configurations compact is to use a lattice, which will be discussed in a later publication [14]. A distance function analogous to $\rho[\cdot, \cdot]$ can be defined on a lattice. When the lattice is finite, this has the virtue of making $\mathcal{M}_D$ a compact manifold. Futhermore, the metric tensor is still idempotent, making it possible to define the generalized inverse (7).

It is not easy to calculate the distance between two arbitrary physical configurations. In particular, the variational problem of finding the gauge transformations minimizing $I[A, B; h, f]$ is a nonlinear set of differential equations. In section 7, it is shown that for $D = 2$ and B a pure gauge, these equations are completely integrable.

5 Geodesics and the intrinsic metric in $\mathcal{M}_D$

The geodesics on $\mathcal{M}_D$ were found by Babelon and Viallet [8] using the tangent-space inner product of Singer [5]. They also conjectured that the distance function $\rho[\alpha, \beta]$ is the geodesic distance function between two configurations α and β in $\mathcal{M}_D$. What is interesting is that even before doing any Riemannian geometry (which takes some justification! See section 6) the geodesics can be easily found from the definition of the metric $\rho[\cdot, \cdot]$ in equations (12) and (13) using unsophisticated methods. Furthermore, Babelon and Viallet's conjecture is simple to verify.

In any metric space, there is a second metric, called the **intrinsic** metric $r[\cdot, \cdot]$ defined as the minimal length of a curve (parametrized by a real number) connecting two points [24]. It is not hard to prove that $r[\cdot, \cdot]$ satisfies all the conditions of a metric

space. It is easy to show that for any two configurations α and β, $r[\alpha, \beta] \geq \rho[\alpha, \beta]$. Furthermore, the two metrics are asymptotically equal as the two points coalesce.

What will be shown in this section is that not only is the intrinsic metric $r[\cdot, \cdot]$ bounded above by $\rho[\cdot, \cdot]$, but the two metrics are actually the same (as conjectured in reference [8]). Thus the distance defined in section 4 between two gauge configurations is the length of a minimal curve joining them together. This curve corresponds to a particular straight line in $\mathcal{U}$ (though it is certainly not straight in $\mathcal{M}_D$).

Only minimal geodesics bewteen pairs of points will be considered in this section.

The intrinsic metric between two points, α and α', is obtained by

1. picking a real number $\epsilon > 0$,
2. considering sets of n points, α_1,...,$\alpha_n \in \mathcal{M}_D$, such that $\rho[\alpha, \alpha_1]$, $\rho[\alpha_1, \alpha_2]$,...$\rho[\alpha_n, \alpha']$ are all less than ϵ,
3. finding the greatest lower bound of the sum of distances $\rho[\alpha, \alpha_1] + \rho[\alpha_1, \alpha_2] + ... + \rho[\alpha_n, \alpha']$ with respect to α_1,...,α_n,
4. finding the greatest lower bound of this sum with respect to the number of interpolating points n and
5. taking the limit as ϵ goes to zero.

This definition can be shown to be equivalent to that of Aleksandroff and Zalgaller [24]. The number of interpolating points, n, generally (but not always) becomes infinite. If a parameter $t \in [0, 1]$ is introduced by $\alpha_j = \alpha(t)$, with $t = j/n$, the points form a minimal curve.

Let A' be a connection in α^r Let $\mathcal{M}_D^{\epsilon,n}$ be the set of n-tuples $(\alpha_1, ...\alpha_n)$ in $\mathcal{M}_D$ such that $\rho[\alpha, \alpha_1] < \epsilon$, $\rho[\alpha_1, \alpha_2] < \epsilon$... $\rho[\alpha_n, \alpha^r] < \epsilon$. Let ${\mathcal{U}_{h,A'}}^{\epsilon,n}$ be the set of n-tuples $(A_1, ...A_n)$ in $\mathcal{U}$ such that $\|A^h - A_1\| < \epsilon$, $\|A_1 - A_2\| < \epsilon$,..., $\|A_n - A'\| < \epsilon$. The definition of the intrinsic metric is

$$\begin{aligned} r[\alpha, \alpha'] &= \lim_{\epsilon \to 0} \inf_{n=0,1,2...} \inf_{(\alpha_1,...\alpha_n) \in {\mathcal{M}_D}^{\epsilon,n}} \{\rho[\alpha, \alpha_1] + \rho[\alpha_1, \alpha_2] \\ &+ \cdots + \rho[\alpha_{n-1}, \alpha_n] + \rho[\alpha_n, \alpha']\} \, . \end{aligned}$$

By virtue of property (14), this is equal to

$$\begin{aligned} r[\alpha, \alpha'] &= \lim_{\epsilon \to 0} \inf_{n=0,1,2...} \inf_{h,h_1,...,h_n \in G_D} \inf_{A \in \alpha, A' \in \alpha'} \inf_{({A_1}^{h_1},...,{A_n}^{h_n}) \in {\mathcal{U}_{h,A'}}^{\epsilon,n}} \{\sqrt{I[A, A_1; h, h_1]} \\ &+ \sqrt{I[A_1, A_2; h_1, h_2]} + \cdots + \sqrt{I[A_{n-1}, A_n; h_{n-1}, h_n]} + \sqrt{I[A_n, A'; h_n, 1]} \ \} \, . \end{aligned}$$

or, equivalently

$$\begin{aligned} r[\alpha, \alpha'] &= \lim_{\epsilon \to 0} \inf_{n=0,1,2...} \inf_{h,h_1,...,h_n \in G_D} \inf_{A \in \alpha, A' \in \alpha', (A_1^{h_1},...,A_n^{h_n}) \in {\mathcal{U}_{h,A'}}^{\epsilon,n}} \{\|A^h - A_1^{h_1}\| \\ &+ \|A_1^{h_1} - A_2^{h_2}\| + \cdots \|A_{n-1}^{h_{n-1}} - A_n^{h_n}\| + \|A_n^{h_n} - A'\|\} \, . \end{aligned} \tag{20}$$

The connections $A_j^{h_j}$ in (20) can be replaced by just A_j, which gives

$$\begin{aligned} r[\alpha,\alpha'] &= \lim_{\epsilon\to 0} \inf_{n=0,1,2\ldots} \inf_{h\in G_D} \inf_{A\in\alpha, A'\in\alpha'} \inf_{A_1,\ldots,A_n\in\mathcal{U}_{h,A'}{}^{\epsilon,n}} \{\|A^h - A_1\| + \|A_1 - A_2\| \\ &+ \cdots + \|A_{n-1} - A_n\| + \|A_n - A'\|\} \ . \end{aligned} \tag{21}$$

Since $\mathcal{U}$ is a linear space, the infimum over A_1,...,A_n of the quantity in brackets {} in (21) is just $\|A^h - A'\|$. Thus (21) reduces to

$$r[\alpha,\alpha'] = \rho[\alpha,\alpha'] \ ,$$

which is what was conjectured by Babelon and Viallet [8].

It is now obvious what the geodesic joining α and β is. Let Ξ be the mapping from $\mathcal{U}$ to $\mathcal{M}_D$ which takes any gauge field configuration into the physical configuration containing that gauge field configuration (for example, $\Xi(A) = \alpha$). If $A \in \alpha$ then there is a choice of $B \in \beta$, which minimizes $\|A - B\|$. The geodesic is the image under Ξ of the line segment in $\mathcal{U}$

$$A(t) = A + Et \ , \tag{22}$$

where $A + E = B$ and $0 \le t \le 1$. This solution was found using different methods by Babelon and Viallet [8]. It is just a straight line in connection space $\mathcal{U}$. However, the geodesic is *not* a straight line in $\mathcal{M}_D$. The geodesic curve is a curve in the space of equivalence classes, $\Xi(A(t))$, and not a curve in the space of connections, $A(t)$. Notice that because the geodesic is parametrized by a straight line in connection space, $\mathcal{U}$, it cannot contain a point conjugate to the point $\Xi(A)$ at $t = 0$. In fact, one of the technical inconveniences of the formalism used here is that conjugate points and closed geodesics cannot be obtained from just one solution (22). They must be made by more than one such solution (22) pieced together at the ends by gauge transformations.

Other extremal curves, conjugate points and closed geodesics can probably be examined using the methods developed in this paper. It seems reasonable that the curve $\Xi(A(t))$ is still a geodesic for $t < 0$ and $t > 1$. Does extending the geodesic in this way always make a closed curve? If so, there are an infinite number of Gribov copies of A along the line $A(t) \in \mathcal{U}$ [22].

6 The singular metric tensor on $\mathcal{M}_D$

Rather picking out members of each equivalence class of gauge fields α by than gauge-fixing, the philosophy advocated in this paper is to work directly with the space of equivalence classes. A straightforward way to do this is to use connections in $\mathcal{U}$ as the coordinates of equivalence classes in $\mathcal{M}_D$. The advantage of such a procedure is that the Gribov problem of copies is never confronted [22]. The disadvantage is that the metric tensor is singular. Here the word "singular" means that some eigenvalues of the metric tensor are zero; it does not mean that the metric is discontinuous or divergent at some point of configuration space $\mathcal{M}_D$. In spite of this singularity, it is possible to define standard geometric quantities on $\mathcal{M}_D$, such as the the volume element, the

Laplacian and the Riemann curvature tensor. Since differential calculus with singular metric tensors does not seem to be discussed in textbooks on geometry, a review is given in the appendix.

To obtain the metric tensor, one would like to consider two "close" physical configurations α, β containing gauge configurations A and $A+\delta A$, respectively, where δA is infinitesmal, thereafter expanding to quadratic order in δA. If this procedure is to make sense, certain criteria must be satisfied:

1. If β is some configuration near α and α contains A, then β contains a gauge field B of the form $A+\delta A$.

2. The connections in $\mathcal{U}$ are mapped continuously to physical configurations in $\mathcal{M}_D$.

3. If β contains two distinct elements $B = A+\delta A$ and $C = A + (\delta A)'$ then it must be so that B and C differ by a "small" gauge transformation $\mathcal{D}^A\omega \approx (\delta A)' - \delta A$.

Criteria 1 and 2 are needed in order to be able to use connection space, $\mathcal{U}$, as coordinates for physical configuration space, $\mathcal{M}_D$. Criterion 3 is required if the metric is to be Taylor expanded in these coordinates to obtain the metric tensor and its derivatives. One might worry that criterion 3 is false; in particular that C could be a Gribov copy of B. This would mean that B and C are equivalent to each other by a "large" gauge transformation g (of Chern-Simons number zero for $D=3$), $B^g = C$, but are not equivalent by a "small" gauge transformation, $\exp -i\omega(x)$, for which $\mathcal{D}^A\omega \approx C - B$. Fortunately, all the criteria are true, as will be shown below.

Before proving these criteria, a few definitions will be made. Recall from section 4 that $\mathcal{U}$ is a metric space, with the metric $\|\cdot - \cdot\|$. Recall from section 5 that Ξ is the mapping from $\mathcal{U}$ to $\mathcal{M}_D$ which takes any gauge field configuration into the physical configuration containing that gauge field configuration. The open ball $s_r(\alpha)$ of radius r around α is defined by

$$s_r(\alpha) = \{\gamma : \rho[\gamma,\alpha] < r\} .$$

and the open ball $S_r(A)$ of radius r around A is defined by

$$S_r(A) = \{C : \|A - C\| < r\} .$$

Criteria 1, 2 and 3, respectively, can now be stated as:

Proposition 6.1 *: Suppose $A \in \mathcal{U}$ is any gauge field configuration and $\Xi(A) = \alpha$. Then $s_r(\alpha)$ is the image under Ξ of $S_r(A)$, or*

$$\Xi(S_r(A)) = s_r(\alpha) . \tag{23}$$

Proof: By the definition of $\rho[\cdot,\cdot]$ and $\|\ \|$ for any $B \in S_r(A)$,

$$\rho[\alpha,\Xi(B)] = \rho[\Xi(A),\Xi(B)] \leq \sqrt{I[A,B;1,1]} = \|A - B\| .$$

Therefore $\Xi(S_r(A))$is contained in $s_r(\alpha)$. Now suppose β is in $s_r(\alpha)$. That means, that there is a $B \in \beta$ such that there is sequence of gauge transformations h_1, h_2,..., with $\|\lim B^{h_n} - A\| < r$. But then, for sufficiently large n, $\|B^{h_n} - A\| < r$. Therefore, since $B^{h_n} \in \beta$ and $B^{h_n} \in S_r(A)$, $\Xi(S_r(A))$ is contained in $s_r(\alpha)$. Equation (23) follows.

Proposition 6.2 *: The mapping Ξ is continuous.*

Proof: In fact, a stronger condition can be proved, namely that Ξ is uniformly continuous. This means that for any $\epsilon > 0$ there exists a $\delta > 0$ such that if $\|A - B\| < \delta$ then $\rho[\Xi(A), \Xi(B)] < \epsilon$. For the case at hand, it is obvious that one can choose $\delta = \epsilon$.

Proposition 6.3 *: Suppose $B, C \in S_r(A)$ and that $\Xi(B) = \Xi(C) = \beta$. For r sufficiently small, there exists a gauge transformation $g(x) \in SU(N)$ such that*

$$C_k(x) = g(x)^{-1} B_k(x) g(x) + i g(x)^{-1} \partial_k g(x) \;, \tag{24}$$

and such that g can be expanded in the form:

$$\begin{aligned} g(x) = 1 - i \int_M d^D x \sum_{k=1}^{D} S_k(x)[C_k(x) - B_k(x)] T_k(x) \\ \\ - \int_M d^D x \int_M d^D y \sum_{k=1}^{D} \sum_{l=1}^{D} U_{k,l}(x,y)[C_k(x) - B_k(x)] \\ \times V_{k,l}(x,y)[C_l(y) - B_l(y)] W_{k,l}(x,y) + O(r)^3 \;, \end{aligned} \tag{25}$$

where $S_k(x)$, $T_k(x)$, $U_{k,l}(x,y)$, $V_{k,l}(x,y)$ and $W_{k,l}(x,y)$ are all matrix valued generalized functions.

Proof: Equation (24) can be solved explicitly for $g(x)$, given that B and C are gauge-equivalent. For each point, $x \in M$, define l to be the straight-line path from the origin to x. Consider the expression

$$g(x) = [\mathcal{P} \exp -i \int_0^x dz \cdot B(z)] \; g(0) \; [\mathcal{P} \exp -i \int_0^x dz \cdot C(z)]^\dagger \;, \tag{26}$$

where both integrations are along l and where $g(0) \in SU(N)$ is an integration constant. Since any solution of the form (26) formally solves (24), $g(0)$ can be chosen to be unity. The task is then to show that (26) really exists as a gauge transformation. Now (26) can be functionally expanded in $C - B$ as

$$\begin{aligned} g(x) = 1 \;\; &+ \;\; i \sum_k \int_0^x dy^k [\mathcal{P} \exp -i \int_y^x dz \cdot B(z)][C_k(y) - B_k(y)] \\ &\times \;\; [\mathcal{P} \exp -i \int_y^x dz \cdot C(z)]^\dagger \\ &+ (\textit{quadratic term in } \; C - B) + O(C - B)^3 \;. \end{aligned} \tag{27}$$

Since $\|C - B\| < r$, making r sufficiently small (recall that M is compact) guarantees that this expansion converges and that it may be differentiated term-by-term with respect to x. One can also check that it satisfies (10). Note that equation (27) is of the form (25). This concludes the proof.

The rigorous part of this paper is now finished. What remains in this section could presumably be made rigorous by being careful with domains of covariant derivative operators on the Hilbert space $\mathcal{U}$.

Now that the use of gauge fields as local coordinates on the space of physical configurations is justified, the metric tensor can be obtained by means of a functional Taylor expansion $g \approx 1 - i\omega$. Suppose that the distance between α and β in $\mathcal{M}_D$ is a small quantity $\rho[\alpha, \beta] = d\rho$, and the distance in $\mathcal{U}$ between A in α and B in β, namely $\|B - A\|$, is also small. Then if δA is defined as $B - A$,

$$d\rho^2 = \min\{ \; \frac{1}{2} \int_M d^D x \; tr \; \sum_{j=1}^{D} [A_j^g - A_j - \delta A_j]^2 \;\; : \; g \in G_D \} \; . \tag{28}$$

In (28), Proposition 6.3 has been use to replace the infimum by the minimum. The variational condition on $g \approx 1 - i\omega$ is

$$\mathcal{D}^j \mathcal{D}_j \omega = \mathcal{D}^j \delta A_j \; . \tag{29}$$

The solution of (29) exists by virtue of the arguments given at the end of section 4. It is

$$\omega = \mathcal{P} \frac{1}{\mathcal{D}^2} \, \mathcal{D}^j \delta A_j \; , \tag{30}$$

where in the Green's function $\mathcal{P}\frac{1}{\mathcal{D}^2}$ taking the principle value projects out the zero modes of $\mathcal{D}^2$ according to the prescription (7) in section 3. Thus the problem with non-generic points raised in reference [8] never arises. In fact, it is clear from section 4 that such a problem shouldn't occur since $\mathcal{M}_D$ is a complete metric space. Substituting (30) back into (28) then gives

$$d\rho^2 = [\int_M d^D x \sum_{j=1}^{D} \sum_{a=1}^{N^2-1}] \, [\int_M d^D y \sum_{k=1}^{D} \sum_{b=1}^{N^2-1}] \; G_{(x,j,a)\;(y,k,b)} \; \delta A_j^a(x) \delta A_k^b(y) \; ,$$

where the metric tensor is

$$G_{(x,j,a)\;(y,k,b)} = \delta_{j\,k} \delta_{a\,b} \delta^D(x - y) - (\mathcal{D}_j \; \mathcal{P} \frac{1}{\mathcal{D}^2} \; \mathcal{D}_k)_{a\,b} \; \delta^D(x - y) \; , \tag{31}$$

which is (8). The metric tensor clearly defines a symmetric form. It is probably possible (being careful in defining some dense subspace of Hilbert space) to define it as a sef-adjoint operator. As stated earlier, G is the projection operator onto small variations of A which are not gauge transformations. Therefore G is singular. This singularity is the price to pay for not fixing the gauge.

As already pointed out in section 3, the metric tensor should be intepreted as its own inverse. This tensor needs to be regularized in the kinetic term of the Yang-Mills Hamiltonian.

The analysis above can be repeated for an Abelian gauge theory. The result for the metric tensor is not very different from (31):

$$G_{(x,j)\;(y,k)} = \delta_{j\,k} \delta^D(x - y) - \partial_j \; \mathcal{P} \frac{1}{\partial^2} \; \partial_k \; \delta^D(x - y) \; .$$

Since this metric tensor is independent of the gauge field, $\mathcal{M}_D$ is flat. The metric is easy to calculate for arbitrarily (not infinitesmally) seperated Abelian gauge fields $A_k(x)$ and $B_k(x)$ in equivalence classes α and β, respectively. For $D = 2$, it is

$$\rho[\alpha, \beta] = -\int d^2x \int d^2y \, [\nabla\times(\mathbf{A}(\mathbf{x}) - \mathbf{B}(\mathbf{x}))] \, \frac{1}{2\pi} \, \ln \frac{|x-y|}{a} \, [\nabla\times(\mathbf{A}(\mathbf{y}) - \mathbf{B}(\mathbf{y}))] \; ,$$

where a has dimensions of length, and for $D = 3$ it is

$$\rho[\alpha, \beta] = \int d^3x \int d^3y \, [\nabla\times(\mathbf{A}(\mathbf{x}) - \mathbf{B}(\mathbf{x}))] \cdot \frac{1}{4\pi|x-y|} \, [\nabla\times(\mathbf{A}(\mathbf{y}) - \mathbf{B}(\mathbf{y}))] \; .$$

The measure of integration on $\mathcal{M}_D$ needs to be defined to make the space of states of the Yang-Mills theory meaningful. Formally, if two wave functionals are $\Psi[\alpha]$ and $\Phi[\alpha]$ and Q is some operator, the matrix element is the functional integral

$$< \Psi|Q|\Phi >= \frac{\int dA \, \sqrt{det'G} \; \Psi[\Xi(A)]^* \; Q \; \Phi[\Xi(A)]}{\int dA \, \sqrt{det'G} \; \Psi[\Xi(A)]^* \; \Phi[\Xi(A)]} \; , \qquad (32)$$

where $\sqrt{det'G}$ is the determinant of the metric tensor after the removal of zero modes. The determinant $\sqrt{det'G}$ is unity in the unregulated theory (it is still unity with lattice regularization [14]). However, the integrals in the numerator and denominator of (32) are divergent unless a regulator is introduced.

There is no difficulty with the expression (32) if the volume of each gauge orbit $O_\alpha = \{A \in \mathcal{U} \; : \Xi(A) = \alpha\}$ is independent of α. However, this is not true everywhere on $\mathcal{M}_D$. At certain points of $\mathcal{M}_D$, called reducible connections in reference [19], there is a nontrivial invariant subgroup of gauge transformations which preserve the gauge configuration. At these points, the gauge orbit has zero volume (if the invariant subgroup is infinite) or a nonzero, but reduced volume (if the invariant subgroup is finite. The volume must be divided by the order of this subgroup). At a reducible connection A, there is a special-unitary-matrix-valued function $g(x) \in SU(N)$ which is an eigenfunction of $\mathcal{D}^2$, with unit eigenvalue, i.e.

$$\mathcal{D}^2 g = g \; .$$

Under a small perturbation of A, such a property is not generally maintained. Therefore (while I have not studied this question carefully) the reducible connections are a set of measure zero in the integral (32) and should present no fundamental difficulty.

7 An integrable system related to the metric of 2+1-dimensional gauge theory

In section 4 the distance between arbitrary two Yang-Mills configurations in $\mathcal{M}_D$ was defined and proven to be a metric. Before pursuing the infinitesmal form of the metric further, a good question is: how close one can come to actually calculating the metric $\rho[\cdot, \cdot]$ for two arbitrary points in $\mathcal{M}_D$? The answer, at least for the time being is: not

very. However, the problem has some interesting aspects which may eventually point toward a solution.

The first step towards a general calculation is to minimize $I[A, B; h]$. If this functional is minimized, then for $N^2 - 1$ real numbers, $\omega^a(x)$ at each point x.

$$\frac{\delta I[A, B; he^{i\omega^a t_a}]}{\delta \omega^a(x)} = 0 \; . \tag{33}$$

For the case of B equal to a pure gauge, this problem has been considered before, because its solution would make it possible to locate the fundamental region [15, 25].

The variational action principle is that of a classical chiral sigma model, implying the differential equations in D dimensions

$$\partial^j [(A_j^h(x))^a - B_j^a(x)] - f_{a\,b\,c}(A_j^h(x))^b B_j^c(x) = 0 \; , \tag{34}$$

where

$$A_j^h(x) = ih(x)^{-1}\partial_j h(x) + h(x)^{-1} A_j(x) h(x) \; . \tag{35}$$

The usual chiral sigma model, described by (34) with both $A_k(x) = 0$ and $B_k(x) = 0$, is well known to be integrable with an infinite set of conserved charges [26].

The condition (33) only guarantees that $I[A, B; h]$ is an extremum, which may not even be a local minimum. Notice that $I[A, B; h]$ is the energy of a gauged spin model with a fixed external gauge field; in other words a spin glass. The specific solution needed for calculating $\rho[\alpha, \beta]$ is the *absolute* minimum of $I[A, B; h]$ [15, 25]. The absolute minimum of the energy of a spin glass is a notoriously difficult quantity to determine [27].

Notice that if $B = 0$, (33) becomes

$$\partial^j [(A_j^h(x)]^a = 0 \; , \tag{36}$$

which is the Coulomb gauge condition [15, 16, 17]. The solution of (36) is certainly not unique [22]. Nonetheless, the analysis presented in this section seems worthwhile; eventually it may help to shed light on the nature of the absolute minimum.

In $2 + 1$ dimensions, the two-dimensional nonlinear equations (34) and (35) are completely integrable when $B_j(x) = 0$; that is they may be written as the compatibility conditions of a two sets of linear equations. This means that the these equations may be expressed as the vanishing of the commutator of two operators, called a Lax pair. A reason why it would be useful to obtain the solution of these equations with $B_j(x) = 0$, in order to find $\rho[\alpha, 0]$, where 0 is the equivalence class of pure gauge configurations, is that this quantity yields a bound on the distance between any two physical configurations α and β in $\mathcal{M}_2$, by the triangle inquality (18)

$$\rho[\alpha, 0] + \rho[\beta, 0] \;\geq\; \rho[\alpha, \beta] \; .$$

Furthermore, it happens that the value of the metric when one of the points is a pure gauge configuration is one of the most physically interesting cases, as will be discussed in sections 10 and 11.

There is another way of writing (34) and (35) for h when $B_j(x) = 0$. Define the curvature free gauge field H by

$$H_j(x) = ih(x)^{-1}\partial_j h(x) \ , \tag{37}$$

so that for any contour from a boundary point y to the point x

$$h(x) = \mathcal{P}\exp -i\int_y^x dz \cdot H(z) \equiv V_y^x[H] \ .$$

The equations (34), (35) and (37) are therefore the same as

$$\begin{aligned} [\partial_1 - iH_1(x), \partial_2 - iH_2(x)] &= 0 \ , \\ \partial^k\{H_k(x) + (V_y^x[H])^\dagger A_k(x) V_y^x[H]\} &= 0 \ . \end{aligned} \tag{38}$$

Define $G(x;[A],[H])$ by

$$G(x;[A],[H]) = \frac{1}{\partial_1 - i[H_1(x),\cdot] + \lambda\partial_2}\, \partial^k\{(V_y^x[H])^\dagger A_k(x) V_y^x[H]\} \ , \tag{39}$$

with boundary condition $G(x;[A],[H]) \to 0$ as $x^1 \to \pm\infty$. It is not generally possible to integrate (39) in closed form for $G(x;[A],[H])$, but this quantity can be defined as a formal power series. Equation (38) is equivalent to

$$[\partial_1 - iH_1(x) - \lambda\partial_2,\ \partial_2 - iH_2(x) + \lambda\partial_1 - i\lambda G(x;[A],[H])] = 0 \ , \tag{40}$$

for arbitrary parameter λ. The two operators in the commutator in the left-hand-side of (40) is the Lax pair.

If the coordinate x^2 is identified with the time and if the gauge field $H_k(x)$ falls to zero as $x^1 \to \pm\infty$, equation (40) can be used to derive an infinite tower of conserved charges [26]. Define the connection $K_j(x) \in su(N)$ by

$$K_1 = H_1 \ , \quad K_2 = H_2 + G(x;[A],[H]) \ . \tag{41}$$

Upon substitution of (41), the equations of motion, (40), take the form

$$[\partial_1 - iK_1(x) - \lambda\partial_2,\ \partial_2 - iK_2(x) + \lambda\partial_1] = 0 \ .$$

By making a rotation of angle $\theta = \tan^{-1}\lambda$, (24) becomes

$$[\partial_1 - i\frac{1}{1+\lambda^2}(K_1 + \lambda K_2),\ \partial_2 - i\frac{1}{1+\lambda^2}(K_2 - \lambda K_1)] = 0 \ .$$

Thus

$$K_1' = \frac{1}{1+\lambda^2}(K_1 + \lambda K_2) \ , \quad K_2' = \frac{1}{1+\lambda^2}(K_2 - \lambda K_1)$$

is a gauge connection with vanishing curvature. Consider now an oriented rectangle R constructed by connecting in order the four points $x_a = (x_a^1, x_a^2) = (-L, t)$, $x_b = (L, t)$,

$x_c = (L, t + \delta t)$ and $x_d = (-L, t + \delta t)$. Then the path-ordered exponential of the line integral of K'_j from x_a to x_b to x_c to x_d and finally back to x_a is equal to unity, i.e.

$$\mathcal{P} \exp -i \int_R dz \cdot K'(z) = 1 \; . \tag{42}$$

By the boundary condition on G and (42), the path-ordered exponential of the line integral of K'_1 on the rectangular path from $x_a = (L, t)$ to $x_b = (-L, t)$ must be independent of t as $L \to \infty$. Thus the expression

$$Q(\lambda) = \mathcal{P} \exp -i \int_{-\infty}^{\infty} dx^1 \cdot \frac{1}{1+\lambda^2}(K_1 + \lambda K_2) \tag{43}$$

is a constant (that is, independent of $t = x^2$) for all λ. The coefficients of the Taylor expansion of $Q(\lambda)$ in λ are the conserved charges.

8 Affine connections and the curvature of $\mathcal{M}_D$

With the metric tensor found in section 6, geometric quantities describing the manifold, such as the Riemann tensor can be calculated [5, 8].

To make the discussion simpler, the curvature will first be considered away from points of $\mathcal{M}_D$ where $\mathcal{D}^2$ has zero modes. The principle value prescription can then be ignored in (31) and

$$G_{(x,j,a)\;(y,k,b)} = \delta_{j\;k}\delta_{a\;b}\delta^D(x-y) - (\mathcal{D}_j \frac{1}{\mathcal{D}^2}\; \mathcal{D}_k)_{a\;b}\; \delta^D(x-y) \; . \tag{44}$$

The results then agree with references [5, 8].

The derivative of the metric tensor (44) with respect to a coordinate is

$$\begin{aligned} \frac{\delta}{\delta\; A^l_c(z)} G_{(x,j,a)\;(y,k,b)} &= G_{(x,j,a)\;(z,l,d)}(f_c \frac{1}{\mathcal{D}^2_y}\, \mathcal{D}_{y\;k})_{d\;b}\delta^D(x-y) \\ &- (\mathcal{D}_{x\;j}\,\frac{1}{\mathcal{D}^2_x} f_c)_{a\;d}\delta^D(x-z)\; G_{(z,j,d)\;(y,k,b)} \;, \end{aligned}$$

where $(f_c)ab \equiv f^{acb}$.

Some simple identities (all of which can be proved by integrating over a distribution and using integration by parts) which are used implicitly in the remainder of this section are

$$\mathcal{D}_{x,j}\frac{1}{\mathcal{D}^2_x}\mathcal{D}_{x,k}\delta^D(x-y) = \mathcal{D}_{y,j}\frac{1}{\mathcal{D}^2_y}\mathcal{D}_{y,k}\delta^D(x-y) \;,$$

$$\frac{1}{\mathcal{D}^2_x}\mathcal{D}_{x,k}\delta^D(x-y) = -\frac{1}{\mathcal{D}^2_y}\mathcal{D}_{y,k}\delta^D(x-y) \;,$$

$$\mathcal{D}_{x,j}\frac{1}{\mathcal{D}^2_x}\delta^D(x-y) = -\mathcal{D}_{y,j}\frac{1}{\mathcal{D}^2_y}\delta^D(x-y) \;,$$

and

$$G_{(x,j,a)\;(y,k,b)} = G_{(y,j,a)\;(x,k,b)}\ .$$

The triple "indices" such as (x,j,a), (y,k,b), (z,l,c) and (u,m,g) will sometimes (though not always) be written as X, Y, Z and U respectively, and such capital Roman letters will be used generally to refer to a triple consisting of a point of M, a coordinate index and a group index. For example, the notation for a tensor (or other) component $S^X_{Y\,Z}$ means $S^{(x,j,a)}_{(y,k,b)\;(z,l,c)}$. The Einstein summation convention will be used for capital Roman letters, including integration over M, as well as for the discrete coordinate and group indices. However integration over M will not be assumed when coordinates x, y, z, etc. appear more than once in an expression. When using capital Roman indices, functional derivatives will be written as partial derivatives, e.g. $\frac{\delta}{\delta A^a_j(x)} = \partial_X$. The components of the inverse metric tensor $(G^{-1})^{U\,V}$ (which, as shown earlier is equal to $G_{U\,V}$) will be denoted $G^{U\,V}$, and $(G^{-1})^{U\,Z}G_{Z\,V}$ will be written as G^V_U.

Riemannian manifolds with singular metric tensors are discussed in the appendix. It is shown there that a multi-index object is a tensor only if it is unaffected by projecting on any of its indices, e.g.

$$C^{X\;Y\;S}_Z G^Z_Q = C^{X\;Y\;S}_Q\ ,\ \ C^{Z\;Y\;S}_Q G^X_Z = C^{X\;Y\;S}_Q\ ,$$

etc. The covariant derivative must also have this property for each of its indices.

From the discussion in the appendix, it follows that the covariant functional derivative on tensors with upper indices is

$$\begin{aligned}\nabla_X A^{U_1\,\ldots\,U_M} &= \partial_X A^{U_1\,\ldots\,U_M} \\ &+ \Omega^{U_1}_{XV_1}\,G^{U_2}{}_{V_2}\cdots G^{U_M}{}_{V_M}A^{V_1\,\ldots\,V_M} + \cdots \\ &+ G^{U_1}{}_{V_1}\cdots G^{U_{M-1}}{}_{V_{M-1}}\Omega^{U_M}_{XV_M}\,A^{V_1\,\ldots\,V_M}\ .\end{aligned}$$

and that on tensors with lower indices is

$$\begin{aligned}\nabla_X A_{U_1\,\ldots\,U_M} &= \partial_X A_{U_1\,\ldots\,U_M} \\ &- \Gamma^{Z_1}_{U_1 X}\,G^{Z_2}{}_{U_2}\cdots G^{Z_M}{}_{U_M}A_{Z_1\,\ldots\,Z_M} - \cdots \\ &- G^{Z_1}{}_{U_1}\cdots g^{Z_{M-1}}{}_{U_{M-1}}\Gamma^{Z_M}_{U_M X}\,A_{Z_1\,\ldots\,Z_M}\ ,\end{aligned}$$

where there are two types of affine connections, Ω and Γ. The latter have the form of Christoffel symbols,

$$\Gamma^U_{X\,Y} = G^{U\,Z}\Gamma_{Z\,X\,Y}\ ,$$

where

$$\begin{aligned}\Gamma_{Z\,X\,Y} &= \frac{1}{2}(\partial_X G_{Z\,Y} + \partial_Y G_{Z\,X} - \partial_Z G_{X\,Y}) \\ &= \frac{1}{2}[\frac{\delta}{\delta A^a_j(x)}G_{(z,l,c)\;(y,k,b)} + \frac{\delta}{\delta A^b_k(y)}G_{(z,l,c)\;(x,j,a)} - \frac{\delta}{\delta A^c_l(z)}G_{(x,j,a)\;(y,k,b)}]\ ,\end{aligned} \tag{45}$$

while the former are given by

$$\Omega^U_{VX} = -\partial_X G^U{}_V + \Gamma^U_{VX} \; .$$

The functional derivative of the metric tensor with respect to a coordinate is

$$\begin{aligned} \frac{\delta}{\delta A^l_c(z)} G_{(x,j,a)\;(y,k,b)} &= G_{(x,j,a)\;(z,l,d)} (f_c \; \frac{1}{\mathcal{D}^2_y} \mathcal{D}_{y\;k})_{d\;b} \delta^D(x-y) \\ &- \; (\mathcal{D}_{x\;j} \; \frac{1}{\mathcal{D}^2_x} f_c)_{a\;d} \delta^D(x-z) \; G_{(z,j,d)\;(y,k,b)} \; . \end{aligned}$$

The reader should keep in mind that these expressions are not valid on all points of $\mathcal{M}_D$; however, they are valid on the dense set of points for which the spectrum of $\frac{1}{\mathcal{D}^2}$ is well-defined. Evaluating (45) yields

$$\begin{aligned} \Gamma_{Z\;X\;Y} &= \frac{1}{2} G_{(z,l,c)(x,j,f)} f^{afd} (\frac{1}{\mathcal{D}^2_z} \mathcal{D}_{z,k})_{db} \delta^D(z-y) \\ &+ \frac{1}{2} f^{afd} (\mathcal{D}_{z,l} \frac{1}{\mathcal{D}^2_z})_{cd} \delta^D(x-z) G_{(x,j,e)(y,k,b)} \\ &+ \frac{1}{2} G_{(z,l,c)(y,k,f)} f^{bfd} (\frac{1}{\mathcal{D}^2_z} \mathcal{D}_{z,j})_{da} \delta^D(z-x) \\ &+ \frac{1}{2} f^{cfd} (\mathcal{D}_{z,l} \frac{1}{\mathcal{D}^2_z})_{cd} \delta^D(y-z) G_{(y,k,e)(x,j,a)} \\ &- \frac{1}{2} G_{(x,j,a)(z,l,f)} f^{cfd} (\frac{1}{\mathcal{D}^2_x} \mathcal{D}_{x,j})_{db} \delta^D(x-y) \\ &- \frac{1}{2} f^{cde} (\mathcal{D}_{z,l} \frac{1}{\mathcal{D}^2_z})_{ad} \delta^D(x-z) G_{(z,l,e)(y,k,b)} \; . \end{aligned} \tag{46}$$

Notice that no derivatives act on components of the metric tensor in any of the terms in this expression. Upon contraction with $G^{U\;Z}$ to get $\Gamma^U_{X\;Y}$, the second and fourth terms from (46) vanish. The result is

$$\begin{aligned} \Gamma^U_{X\;Y} &= \frac{1}{2} \int d^D z \; G^{(u,m,g)\;(z,l,c)} G_{(z,l,c)\;(x,j,d)} (f_a \; \frac{1}{\mathcal{D}^2_y} \mathcal{D}_{y,k})_{d\;b} \delta^D(z-y) \\ &+ \frac{1}{2} G^{(u,m,g)}_{(y,k,d)} (f_b \; \frac{1}{\mathcal{D}^2_x} \mathcal{D}_{x,k})_{d\;c} \delta^D(x-y) \\ &- \frac{1}{2} \int d^D z \; G^{(u,m,g)\;(z,l,c)} G_{(z,l,d)\;(x,j,a)} (f_c \; \frac{1}{\mathcal{D}^2_y} \mathcal{D}_{y,k})_{d\;b} \delta^D(x-y) \\ &+ \frac{1}{2} \int d^D z (\mathcal{D}_{x,j} \; \frac{1}{\mathcal{D}^2_x} \; f_c)_{a\;d} \delta^D(x-z) \; G_{(z,j,d)\;(y,k,b)} G^{(u,m,g)\;(z,j,c)} \; . \end{aligned} \tag{47}$$

Notice that

$$\Gamma^X_{X\;Y} = 0 \; ,$$

because contracting the indices in (47) annihilates the second and fourth terms, while the first and third terms cancel. Furthermore, it is straightforward to show that $\partial_X G^Y_Y = 0$ and thus

$$\Omega^Y_{Y\ X} = 0\ .$$

As the reader can check from (47), contracting both of the lower indices of Γ or Ω with the inverse metric gives zero, i.e.

$$\Gamma^X_{U\ Y} G^{U\ Z} G^{Y\ V} = 0\ ,\ \ \Omega^X_{U\ Y} G^{U\ Z} G^{Y\ V} = 0. \tag{48}$$

Furthermore, the partial derivative of $G_{X\ Y}$ also vanishes when projected thus:

$$G^X_Z G^Y_V \partial_U G_{X\ Y} = 0\ . \tag{49}$$

While the proof for manifolds with singular metric tensors is not as simple as the nonsingular case, it is true that the covariant derivative of the metric tensor is zero. This is shown in the appendix.

Some understanding of the details of the spectrum of the kinetic term can be obtained by examining the Ricci curvature [13]. This is defined in the usual way as a contraction of the Riemann tensor. This Riemann tensor was first presented, though not explained for $\mathcal{M}_D$ by Singer[5], and was discussed in some more detail by Babelon and Viallet [8]. Sometimes performing tensor contractions on $\mathcal{M}_D$, which requires setting two points of physical space M equal, requires ultraviolet regularization. This is true of the contractions used in the Laplacian and the Ricci tensor. This point has emphasized by Singer [5]. Singer proposed regularizing this divergence by inserting a power of the covariant Laplacian, $(-\mathcal{D}^2)^{-s}$, or some other gauge-invariant damping factor inside tensor contractions. From the perspective taken here, such a construction, though probably correct, is less natural than beginning with a regularized metric space [14].

The Riemann curvature tensor is shown in the appendix to be

$$R^U_{X\ Y\ Z} = \partial_X \Gamma^U_{Y\ Z} - \partial_Y \Gamma^U_{X\ Z} + \Omega^U_{V\ X}\Omega^V_{Z\ Y} - \Omega^U_{V\ Y}\Omega^V_{Z\ X}\ . \tag{50}$$

Notice that for the Yang-Mills case $R_{U\ X\ Y\ Z} = R^U_{X\ Y\ Z}$.

The expression (50) for the Riemann tensor can be simplified by projecting on the three lower indices and using (48) and (49):

$$\begin{aligned} R^U_{X\ Y\ Z} = G^{X'}_X G^{Y'}_Y G^{Z'}_Z R^U_{X'\ Y'\ Z'} &= G^{X'}_X G^{Y'}_Y G^{Z'}_Z (\partial_{X'} \Gamma^U_{Y'\ Z'} - \partial_{Y'} \Gamma^U_{X'\ Z'}) \\ &= G^{Y'}_Y (\partial_{Y'} G^{X'}_X G^{Z'}_Z) \Gamma^U_{X'\ Z'} - G^{X'}_X (\partial_{X'} G^{Y'}_Y G^{Z'}_Z) \Gamma^U_{Y'\ Z'}\ . \end{aligned} \tag{51}$$

Carrying out the differentiations in (51) gives

$$R^U_{X\ Y\ Z} = \{(\partial_R G^S_Z)(G^R_Y G^T_X - G^R_X G^T_Y) + [G^R_Y(\partial_R G^S_X) - G^R_X(\partial_R G^S_Y)]G^T_Z\}\Gamma^U_{T\ S}\ . \tag{52}$$

Only the first and third terms in the expression (47) for $\Gamma^U_{T\ S}$ survive in (52). Introducing the notation

$$B_{s,p} = \frac{i}{\mathcal{D}^2_s}\mathcal{D}_{s,p}\ ,\ \ B^\dagger{}_{s,p} = \mathcal{D}_{s,p}\frac{i}{\mathcal{D}^2_s}\ ,$$

the Riemann tensor is given, after some more work, by

$$R_{(u,m,g)\ (x,j,a)\ (y,k,b)\ (z,l,c)} = \frac{1}{2}\int d^D r \int d^D s \int d^D t \int d^D v$$

$$\times\, [G_{(t,q,e')(v,i,a')}\ (f_{b'} B_{s,p})_{a'\ c'}\delta^D(t-s) - G_{(v,i,b')(t,q,a')}\ (f_{f'}\ B_{s,p})_{a'\ c'}\delta^D(v-s)]$$

$$\times\, ({B^\dagger}_{s,p})_{c'\ d}\delta^D(s-r)\ G_{(v,i,b')\ (u,m,d)} f_h)_{d\,e}$$

$$\times\, [G_{(s,n,e)(z,l,c)} G^{(r,n,h)}_{(y,k,b)} G^{(t,q,e')}_{(x,j,a)} - G_{(s,n,e)(z,l,c)} G^{(r,n,h)}_{(x,k,a)} G^{(t,q,e')}_{(y,k,b)}$$

$$+\, G_{(s,n,e)(x,j,a)} G^{(r,n,h)}_{(y,k,b)} G^{(t,q,e')}_{(z,l,c)} - G_{(s,n,e)(y,k,b)} G^{(r,n,h)}_{(x,j,a)} G^{(t,q,e')}_{(z,l,c)}]\ , \tag{53}$$

where the primed indices have been introduced so as not to run out of symbols. This will be recalculated at points of $\mathcal{M}_D$ where $\mathcal{D}^2$ has zero modes in another paper [14]. An important question is whether the Riemann tensor diverges at these points.

The problem with contracting the indices of (53) to obtain the Ricci curvature is that this involves taking the trace of operators containing $\mathcal{P}\,1/\mathcal{D}^2$. Such expressions are of course ultraviolet divergent [5]. Presumably a good ultraviolet regularization will render $\mathcal{M}_D$ compact (this is definitly true of the lattice [14]). If a regularization is introduced, the eigenvalues of the Ricci curvature can be used to obtain bounds on the ground-state energy and information about the heat kernel of the strongly-coupled theory (i.e. the kinetic term only of H) [13]. Specifically, one considers a nonzero displacement vector ξ^X, which by definition must satisfy $G^X_Y \xi^Y \neq 0$. Then what is relevant is the ratio of quadratic forms

$$Ric(\xi) = \frac{R_{X\ V}\ \xi^Y\ \xi^U}{G_{W\ Z}\ \xi^W\ \xi^Z}\ .$$

In particular, one would like to know if $Ric(\xi)$ has a minimum and if so, whether the minimum is positive. Singer [5] has argued that with his regularization the answer is affirmative, and concludes that the regularized Laplacian on $\mathcal{M}_D$ should have a pure-point spectrum.

9 The Yang-Mills kinetic term

At last the Laplace operator on $\mathcal{M}_D$ will be worked out with the metric tensor

$$G_{(x,j,a)\ (y,k,b)} = \delta_{j\ k}\delta_{a\ b}\delta^D(x-y) - (\mathcal{D}_j\ \mathcal{P}\frac{1}{\mathcal{D}^2}\ \mathcal{D}_k)_{a\ b}\ \delta^D(x-y)\ ,$$

valid on all of $\mathcal{M}_D$. In the appendix, this Laplacian on wave functionals is shown to be

$$\begin{aligned}\Delta\Psi[\alpha] &= -\frac{1}{\sqrt{det'\ G}}\partial_Y(\sqrt{det'\ G}\ G^{Y\,U}\partial_U\Psi[\alpha]) \\ &+ (\partial_Z G^Z_Y)G^{Y\,U}\partial_U\Psi[\alpha] - \frac{1}{2}(\partial_Y G^Z_Z)G^{Y\,U}\partial_U\Psi[\alpha]\ .\end{aligned} \tag{54}$$

It is obvious that the first term is (9). However, there are two new terms present.

The last term of (54) contains a derivative of G^Z_Z. This quantity has to be examined with some care, since it is the trace of an infinite-dimensional operator and requires regularization. In a formal sense

$$G^Z_Z = \mathbf{1}^Z_Z - (\mathcal{D}\,\mathcal{P}\frac{1}{\mathcal{D}^2}\,\mathcal{D})^Z_Z\ .$$

The first term is

$$\mathbf{1}^Z_Z = (N^2-1)\ D\ Vol(M)\ ,$$

where $Vol(M)$ means the volume of M, and is proportional to the dimension of the space of connections $\mathcal{U}$, while the second term is proportional to the dimension of the space of gauge transformations. Thus G^Z_Z is proportional to the dimension of $\mathcal{M}_D$. Roughly speaking, one expects

$$G^Z_Z = (N^2-1)\ D\ Vol(M)\ \frac{\dim \mathcal{M}_D}{\dim \mathcal{U}} \tag{55}$$

If the trace is evaluated carelessly, the result is (since ϵ is infinitesmal)

$$G^Z_Z = \mathbf{1}^Z_Z - \frac{1}{2}\sum_{\pm}(\mathcal{D}^2\ \frac{1}{\mathcal{D}^2 \pm i\epsilon})^Z_Z = i\ \epsilon\ tr(\mathcal{P}\frac{1}{\mathcal{D}^2}) = 0\ ,$$

which cannot be correct if the theory is regularized. In particular, using a finite lattice [14], the ratio of dimensions in (55) will be some positive number. However, what will be true in any case is that the right-hand side of (55) is independent of the physical configuration in $\mathcal{M}_D$. Thus

$$\partial_Y G^Z_Z = 0\ .$$

The other unusual term in (54) is $G^{Y\,U}(\partial_Z G^Z_Y)\partial_U$. The derivative of the metric tensor with respect to a coordinate on any point of $\mathcal{M}_D$ is

$$\partial_Z G^X_Y = \frac{\delta}{\delta A^l_c(z)} G_{(x,j,a)\ (y,k,b)}$$

$$= G_{(x,j,a)\ (z,l,d)}(f_c\ \mathcal{P}\frac{1}{\mathcal{D}^2_y}\,\mathcal{D}_{y\ k})_{d\ b}\delta^D(x-y) - (\mathcal{D}_{x\ j}\ \mathcal{P}\frac{1}{\mathcal{D}^2_x} f_c)_{a\ d}\delta^D(x-z)\ G_{(z,j,d)\ (y,k,b)}$$

$$-\ \pi^2\{\mathcal{D}_{x\ j}\ \delta(\mathcal{D}^2_x)[\mathcal{D}_{x\ l}\ f_c\ \delta^D(y-z) + f_c\ \delta^D(x-z)\ \mathcal{D}_{y\ l}]\delta(\mathcal{D}^2_y)\mathcal{D}_{y\ k}\}\delta^D(x-y)\ . \tag{56}$$

The factors of π can be understood from the relation

$$\partial_Z \mathcal{P}\frac{1}{\mathcal{D}^2_y} = \frac{1}{2}\sum_{\pm}\partial_Z \frac{1}{\mathcal{D}^2_y \pm i\epsilon} = -\frac{1}{2}\sum_{\pm}\frac{1}{\mathcal{D}^2_y \pm i\epsilon}\ \frac{\delta \mathcal{D}^2_y}{\delta A^l_c(z)}\ \frac{1}{\mathcal{D}^2_y \pm i\epsilon}$$

$$= -\frac{1}{2}\sum_{\pm}[\mathcal{P}\frac{1}{\mathcal{D}^2_y} \mp i\pi\delta(\mathcal{D}^2_y)]\ \frac{\delta \mathcal{D}^2_y}{\delta A^l_c(z)}\ [\mathcal{P}\frac{1}{\mathcal{D}^2_y} \mp i\pi\delta(\mathcal{D}^2_y)]\ .$$

Now if $\partial_Z G^X_Y$ is multiplied by $G^{Y\,U}$, contracting on Y, only the second term of (56) survives. Thus

$$(\partial_Z G^Z_Y) G^{Y\,U} = -\int d^D z \int d^D x\; \delta^D(x-z)(\mathcal{D}_{x\,j}\, \mathcal{P}\frac{1}{\mathcal{D}^2_x} f_c)_{a\,d} \delta^D(x-z)\; G^{(u,k,b)}_{(z,j,d)}\;. \qquad (57)$$

This quantity is singular and requires ultraviolet regularization.

The final result of this section is that the Yang-Mills Hamiltonian is

$$H = \frac{e^2}{2}\Delta + \int_M d^D x\; \frac{1}{4e^2}\, tr\, F_{j\,k}(x)^2\;,$$

where

$$\Delta\Psi[\alpha] \;=\; -\partial_Y(G^{Y\,U}\partial_U\Psi[\alpha]) + (\partial_Z G^Z_Y)G^{Y\,U}\partial_U\Psi[\alpha]\;, \qquad (58)$$

and the second term of (58) can be written more explicitly with the use of (57).

10 Magnetic topography of $\mathcal{M}_D$; scale transformations

The metric properties of the manifold $\mathcal{M}_D$ of configurations are relevant to the spectrum of the kinetic term of the $SU(N)$ Yang-Mills Hamiltonian. In order to understand the spectrum at weak coupling, the potential or magnetic energy needs to be considered as well.

A natural starting point is make a relief map of magnetic energy on $\mathcal{M}_D$. The job of magnetic topography is not easy, because there is no obvious relation between this energy and the distance from a pure gauge. Nonetheless, using some simple scaling arguments, at least a partial map can made.

The most interesting result of this section is an unexpected feature of the potential energy function on $\mathcal{M}_3$ (in $3+1$ dimensions). It is found that the potential energy changes dramatically in response to certain small displacements in configuration space from a pure gauge. Furthermore a configuration whose distance from a pure gauge is arbitrarily large can have a potential energy which is arbitrarily small. For the unregularized $2+1$-dimensional Yang-Mills theory, the potential energy on a sphere of any given radius in $\mathcal{M}_2$, centered around a pure gauge configuration takes all possible values.

Suppose that the manifold of physical space, M, is very large. Consider a rescaling of the coordinates and the connection $A \in \mathcal{U}$ by a real factor s:

$$A_j(x) \longrightarrow sA_j(sx)\;. \qquad (59)$$

This rescaling will take place for A^h, as long as $h(x)$ is redefined by

$$h(x) \longrightarrow h(sx)\;. \qquad (60)$$

Under (59) and (60), the distance from an equivalence class of pure gauges, α_0, transforms as

$$\rho[\Xi(A), \alpha_0] \longrightarrow s^{\frac{2-D}{2}} \rho[\Xi(A), \alpha_0] \; . \tag{61}$$

This follows from the definitions (12) and (13). Note that upon rescaling, the minimum of $\sqrt{I[A, B; h, f]}$ must remain a minimum (it does not become a maximum or other extremum) because $s^{\frac{2-D}{2}} > 0$.

Suppose that $A \in \mathcal{U}$ is a particular configuration of finite potential energy, namely one for which the magnetic field $F_{jk}(x)$ is continuous and differentiable, but decays rapidly to zero for x outside some finite bounded region (such as a sphere). This region will be called the **domain** of the magnetic field. By changing the size of the domain and the magnitude of the magnetic field, the distance from some given pure gauge can be made arbitrarily small (except when regularization effects become important) or large (except when volume effects become important).

The potential energy

$$U[\Xi(A)] = \int_M d^D x \, \frac{1}{4e^2} \, tr \, F_{j\,k}^2(x) \; ,$$

transforms as

$$U[\Xi(A)] \longrightarrow s^{4-D} U[\Xi(A)] \; . \tag{62}$$

The transformation properties (61) and (62) imply that for a chosen gauge connection A in (59) and arbitrary scale factor s

$$U \sim \rho^{-2\frac{4-D}{D-2}} \; . \tag{63}$$

Notice that for $2 < D < 4$ the exponent in (63) is negative. Thus it is possible to have arbitrarily large U for arbitrarily small ρ. Since this is true for any initial choice of $A \in \mathcal{U}$, there are an infinite number of physical configurations close to a pure gauge with arbitrarily large potential energy. This statement has nothing to do with the gauge group, and holds for the Abelian theory as well.

If a regulator is introduced, the scaling properties (61) and (62) are violated; but these properties should still hold as long as the domain has a diameter greater than the size of the regulator, such as a lattice spacing, a. Furthermore, configurations near a pure gauge can have energy of order a, though no greater. There does not appear to be any obstacle to using (59), and a regularization will be implicit in the remainder of the discussion.

For Abelian gauge theories, other rescalings besides (59) and (60) can be considered; these take the form

$$A_j(x) \longrightarrow s^{\phi} A_j(sx) \; , \tag{64}$$

where ϕ is any real number. Gauge invariance can still be maintained under (64). This is not possible for Yang-Mills theories, for which only $\phi = 1$ is permitted. Under (64)

$$\rho[\Xi(A), \alpha_0] \longrightarrow s^{\frac{2\phi-D}{2}} \rho[\Xi(A), \alpha_0] \; ,$$

and

$$U[\Xi(A)] \longrightarrow s^{2\phi+1-D} U[\Xi(A)] \ . \tag{65}$$

Consequently, by choosing ϕ satisfying

$$\frac{D-2}{2} < \phi < \frac{D}{2} \tag{66}$$

it is always possible to make the potential energy arbitrarily small for small s, no matter what the dimension. The configuration then becomes one for which the diameter of the domain is very large but the field strength is very small; a quantum wave functional $\Phi[\alpha]$ whose amplitude is largest near this configuration is a long-wavelength photon. This quantum state must be orthogonal to the vacuum $\Psi_0[\alpha]$ because at least one of the two wave functionals is zero anywhere on $\mathcal{M}_3$. This is true since the distance in configuration space between the large-domain/low-energy configuration and the pure gauge configuration becomes infinite as s is taken to zero. Not only are the two states orthogonal, but the energy eigenvalue of $\Phi[\alpha]$ becomes zero in this limit. This is why electrodynamics (at least without magnetic monopoles) has no mass gap in any dimension.

Is it possible to repeat the argument of the last paragraph for a Yang-Mills theory? Remarkably, the answer is yes in $3+1$ dimensions. Notice that for the inequality (65) to be satisfied with $\phi = 1$, the only integer solution for the dimension of M is $D = 3$. This evidently implies that the excitations of QCD include massless particles!

To understand what is happening in a more detail, consider the following trajectory of points α in $\mathcal{M}_3$ (figure 2). Start from the pure gauge configuration $\alpha_{n,0}$ with a particular Chern-Simons integral $C = n$. Distort the configuration by increasing the eigenvalues of magnetic field strength (which are gauge invariant) to be non-zero in a small domain. Call this configuration α_1. In doing so, the configuration α moves along a short path P_1, directed away from $\alpha_{n,0}$, ending at α_1, while the potential energy $U[\alpha]$ increases very slightly along this path. Since this distortion is made continuously, n changes only very slightly (it need no longer be an integer). Now perform a scale transformation, with s varying from 1 to some very small value. The potential energy decreases under this scale transformation. This has the effect of producing a path P_2 from α_1 to a configuration α_2 (which is far from $\alpha_{n,0}$) all the while keeping the potential energy arbitrarily small. Furthermore, it is easy to show by examining (2) that the Chern-Simons integral $C \approx n[\alpha]$ does not change under the scale transformation (59). If P_1 and P_2 are joined to make $P = P_1 + P_2$, the Chern-Simons integral is nearly unchanged along P. The possibility of identifying α_2 with a pure gauge other than $\alpha_{n,0}$ is thereby ruled out. Thus there is no possibility that α_2 is a pure gauge.

The wave functional $\Phi[\alpha]$ which is large near near α_2 is evidently representative of a long-wavelength gluon in the quantum theory. It is a state which is orthogonal to the vacuum and has arbitrarily small energy, by the arguments used for the Abelian theory. How then can there be a gap in the spectrum? Several proposals are made in the next section.

In $2+1$ dimensions, the situation is very different. There is only one point of configuration space $\mathcal{M}_2$ which is an equivalence class of pure gauge fields, namely $\Xi(0)$.

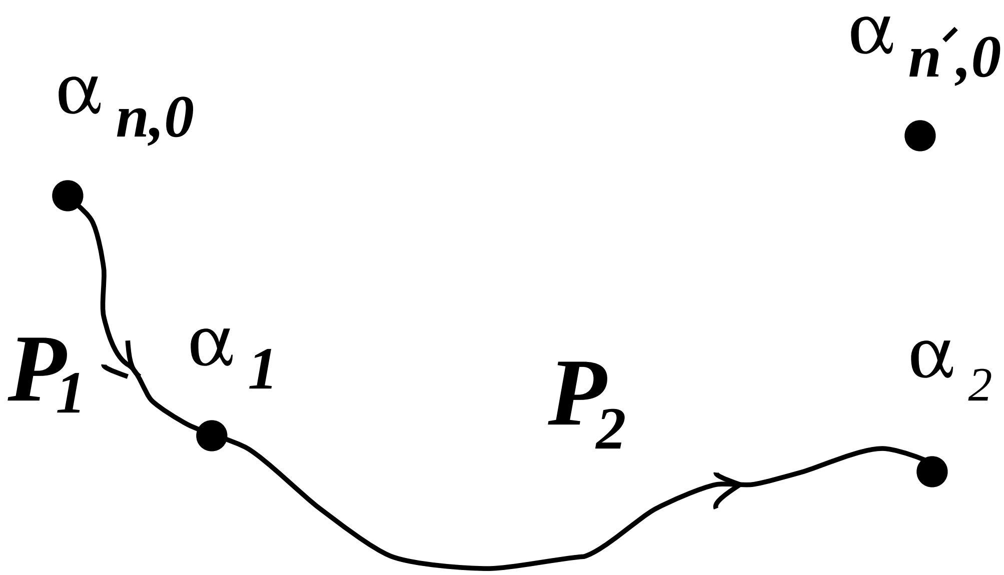

Figure 2: How to make configurations of low potential energy far from a pure gauge in $3+1$-dimensional Yang-Mills theories. The point $\alpha_{n,0}$ is a pure gauge configuration of Chern-Simons number n. The potential energy increases very slightly along the path P_1 from $\alpha_{n,0}$ to α_1. The potential energy decreases along the path P_2 from α_1 to α_2, while the distance from $\alpha_{n,0}$ grows.

Under a rescaling (59), the potential energy of some configuration α is proportional to s^2, but the distance from $\Xi(0)$ does not change at all. The scale transformation moves the configuration around a sphere in $\mathcal{M}_2$, centered around $\Xi(0)$. The range of values of the potential energy on this sphere are all real numbers between zero and infinity.

11 Magnetic topography of $\mathcal{M}_D$; the mass gap

The physical implications of section 10 will now be examined.

Perhaps the first counter-argument against concluding that there is no gap in $3+1$ dimensions which comes to mind is that the potential energy function is extremely "hilly", rising and falling rapidly, on all of $\mathcal{M}_3$ for Yang-Mills theories (as it certainly is near the origin). Then there is the possibility of states which are localized in $\mathcal{M}_3$, as occurs for a random potential, even though there are some directions in which the potential energy falls rapidly. The regions of small potential energy would have large electric (kinetic) energy by the uncertainty principle. In this way, the first excited state could have a finite gap above the ground state. Even if such potential-energy barriers exist, they would all fade away as a scale transformation (59) is made. However,

the uncertainty principle could make the kinetic energy increase under such the scale transformation. Such a mechanism is very close to the picture of extrapolating the effective coupling to large distances [28]; the renormalized potential would then grow under scaling transformations (59) rather than shrink. To investigate this possibility requires exploration of the kinetic energy (the Laplacian on $\mathcal{M}_3$) and the potential energy in a small neighborhood of α_2. It is left as an exercize for the reader to show that formally, i.e. ignoring regularization and renormalization, the Riemann curvature scalar decreases with decreasing s, $R[\alpha] \rightarrow s^{D-2}\, R[\alpha]$, suggesting that curvature effects aren't very important at large scales. Of course, this relation has to be modified by an anomalous dimension. Thus it may be that the quantum theory has a gap to the first excited state, though simple intuition about the classical theory would indicate otherwise. Some examples of quantum-mechanical systems for which this is the case were discussed by Simon in reference [29] (including a model suggested by Goldstone and Jackiw motivated by the Yang-Mills theory).

Another possible explanation for the gap is that the true vacuum is a superposition of states of different Chern-Simons integrals [21]. While there is an energy barrier between α_2 and a pure gauge configuration $\alpha_{n',0}$ of Chern-Simons integral $n' \neq n$, the distance between them, $\rho[\alpha_{n,2}, \alpha_{n',0}]$ could conceivably be very small. If this is the case, it is not possible to create a state orthogonal to the ground state by the arguments above. This is because the quantum state associated with α_2 would have a significant overlap with the true vacuum state, which is not an eigenstate of the Chern-Simons integral. A general argument could be made if an upper bound on $\rho[\alpha_2, \alpha_{n',0}]$ could be found.

In my opinion, both of the mechanisms invoked above for a gap are probably correct for finite N. In the large-N limit [30], the importance of the latter may be diminished, as tunnelings between different Chern-Simons numbers are suppressed [31].

One might also guess that the volume of regions in $\mathcal{M}_3$ far from $\alpha_{n,0}$ is small, making such regions inaccesible. However, this guess is false. To see this, consider some set of points a constant distance from $\alpha_{n,0}$. Under the rescaling, the distance between any pair of points in this set grows at the same rate as the distance between either point in the pair and $\alpha_{n,0}$. Thus spheres in $\mathcal{M}_3$ always increase in size under a rescaling. This is consistent with a decreasing Riemann curvature scalar for decreasing s.

It was shown that for the $2+1$-dimensional Yang-Mills theory, the potential energy on those points in $\mathcal{M}_2$ which are a fixed distance from $\Xi(0)$ takes all possible values. Note that when $D < 2$, the scaling formula (63) indicates that that the potential energy of a configuration *increases* with its distance from $\Xi(0)$ (in fact, for $D = 2 - \epsilon$ the distance decreases as as $s \rightarrow 0$, which is consistent with a gap). It should be true that the distance function $\rho[\cdot, \cdot]$ can be defined in the *renormalized* $2+1$-dimensional Yang-Mills theory (indeed, this was Feynman's viewpoint [4]), and that this property is present there as well. If so, a mass gap should be present. Since the theory is super-renormalizable, it should be much easier to study the renormalized metric and potential energy than in $3+1$ dimensions.

What happens when the space dimension, D, is greater than 3? The potential is constant under a rescaling for $\mathcal{M}_4$, suggesting a massless phase at weak coupling.

However for $D = 5$, the scaling formula again (63) suggests (but does not prove! See below) that the energy of a configuration increases with its distance from a pure gauge. This suggests that the $5+1$-dimensional gauge theory has a mass gap. This theory almost certainly does not confine at weak coupling, which means that it would exhibit a massive Higgs phase [32] when the bare coupling constant is between zero and some value. As far as I know, lattice Monte-Carlo calculations have not ruled out this possibility. To be certain that the potential energy function is convex, it is neccesary to investigate other motions in $\mathcal{M}_5$ besides (59) to see whether that the potential could decrease in other directions.

12 Discussion

What was done was to start at the bottom from analysis and work up to geometry. The result is at least the barest outlines of the space in which physical Yang-Mills configurations live. Through a careful study of the distance function, the kinetic energy operator in the Schrödinger representation was found on this space. The magnetic energy on this space was investigated for different dimensions of space-time.

The point of view taken here was different from that of references [15, 16, 17, 18, 19] in that no gauge fixing (except for the original temporal gauge condition) was imposed. This viewpoint has also been advocated in reference [20]. The drawback of a metric tensor with zero eigenvalues seems outweighed by the advantage of not having to isolate a Gribov domain. To make an analogy with Riemann surfaces, the methods used here and in reference [20] are like the use of automorphic functions, while what the authors of references [15, 16, 17, 18, 19] have been attempting to do is like directly determining the properties of the fundamental region.

It should not be very hard to introduce similar techniques for Euclidean path integrals. Then the analysis here is relevant to the D-dimensional Yang-Mills theory with no gauge fixing, rather that the $D+1$-dimensional Hamiltonian theory with temporal gauge.

There are many issues to examine further. The metric, Laplacian and Ricci curvature need to be carefully recomputed with some explicit regularization [14]. The issue of the measure and reducible connections [19] should also be rexamined on the lattice. Closed geodesics and conjugate points will be very interesting to study [22, 5, 7, 8]. The distance between configurations α_2 and $\alpha_{n',0}$, discussed in section 11, must be estimated. What is most important is that displacements in $\mathcal{M}_D$ other than those discussed in sections 10 and 11 should be investigated.

The $3+1$-dimensional low-energy configuration found in section 10 is not a *perturbative* gluon. The former is a large geodesic distance from a pure gauge in $\mathcal{M}_3$, whereas the latter is a small geodesic distance from a pure gauge.

The metric for the super-renormalized theory in 2+1 dimensions needs to be studied. It may be that it is easier than in $3+1$ dimensions to look at the potential in directions perpendicular to the scale direction studied in section 10. In fact it may be even more straightforward to study the theory in just above two space-time dimensions, i.e. in dimension $(1+\epsilon)+1$ [33], for which the number of degrees of freedom might be small

enough to find a complete picture of the Laplacian and the potential energy function on all of $\mathcal{M}_{1+\epsilon}$.

Appendix: Differential calculus on manifolds with singular metric tensors

For the reader who is uncomfortable with the notion of a singular metric tensor discussed in this paper, the general finite-dimensional case is discussed here. An example is given which describes the distance function on a two-dimensional sphere. Like the metric discussed for Yang-Mills fields, it may be used globally on the manifold; it is not neccesary to introduce additional coordinate charts.

Consider a manifold for which the dimension of tangent space, d, is a fixed number, and for which the number of coordinates, N, of any coordinate chart is also a fixed number. The metric tensor will be singular provided $d < N$. Label the coodinates by $x = (x^1, ..., x^N)$ in some given coordinate chart and label some choice of basis of tangent space at x by $\mathbf{e^1}(x), ...\mathbf{e^d}(x)$. Since the coordinate chart is N-dimensional, there are neccesarily vector fields which are othogonal to tangent space; thus for some functions of coordinates $\phi(x)$

$$\nabla\phi(x) \cdot \mathbf{e^J}(x) = 0,\ J = 1, ..., d\ . \tag{67}$$

In order for (67) to be consistent, a point x in a coordinate chart $\mathcal{C}$ and $x + dx \cdot \nabla\phi(x)$ must be identified. This means that points of the manifold cannot be parametrized by points in $\mathcal{C}$, but rather by curves in $\mathcal{C}$ for $N - d = 1$, by surfaces in $\mathcal{C}$ for $N - d = 2$, etc. Experts in geometry will see that these curves, surfaces, etc. are just fibers of an N-dimensional bundle.

The coordinate-space components of the tangent vector $\mathbf{e}^J(x)$ will be denoted by ${e^J}_j$. A tangent-space vector field can be written $\mathbf{A}(x)$ and has the standard representation

$$\mathbf{A}(x) = \sum_J \mathbf{e}^J(x) A_J(x)\ . \tag{68}$$

Unfortunately, as it stands, (68) is not very useful, because the components $A_J(x)$ are not coordinate components. The latter can be found after first defining an N-dimensional set of vectors $\mathbf{E_1}(x), ...\mathbf{E_N}(x)$; these span a vector space which includes tangent space as a subspace. Then the vector field $\mathbf{A}(x)$ (henceforth called a vector) can be written as

$$\mathbf{A}(x) = \sum_j \mathbf{E_j}(x) A^j(x)\ . \tag{69}$$

However, (69) does not show explicitly that $\mathbf{A}$ is a tangent vector.

For the purposes of both this appendix and the text of this article it will be assumed that

$$\mathbf{E}_i(x) \cdot \mathbf{E}_j(x) = \delta_{ij}\ . \tag{70}$$

In general, one can consider cases where (70) is not so; the only complication is more index clutter, and there is little essential difference. Clearly $e^J{}_j = \mathbf{E}_j \cdot \mathbf{e}^J(x)$.

The metric tensor is defined by

$$g_{ij}(x) = \sum_K e^K{}_i(x) e^K{}_j(x) \; .$$

This is the correct definition because a change in coordinates dx produces a displacement dY in tangent space

$$dY^K = \sum_i e^K{}_i(x) dx^i \; . \tag{71}$$

The arc-length differential ds given by

$$ds^2 = \sum_K dY^K dY^K = g_{ij} \; dx^i dx^j \tag{72}$$

Clearly $g_{ij}(x)$ is singular.

There is a meaningful notion of inverse metric tensor. Consider now another basis of the d-dimensional tangent space, namely the the orthonormal basis of eigenvectors of the matrix $g_{ij}(x)$ with nonzero eigenvalues $\mathbf{S}_1(x), ..., \mathbf{S}_d(x)$, $\mathbf{S}_q(x) \cdot \mathbf{S}_r(x) = \delta_{qr}$. The metric tensor can be written as

$$g_{ij}(x) = \sum_J \lambda_J S^J{}_i(x) S^J{}_j(x) \; , \tag{73}$$

where $\lambda_J \neq 0$. The projection operator into tangent space is

$$P^i{}_j(x) = \sum_J S^J{}_i(x) S^J{}_j(x) \; . \tag{74}$$

The mismatch of upper and lower indices in (74) is not a misprint. The expression for the inverse metric tensor

$$g^{ij} = \sum_J (\lambda_J)^{-1} S^J{}_i S^J{}_j \; , \tag{75}$$

also displays such a mismatch. Notice that

$$P^i{}_j = g^{ik} g_{kj} \equiv g^i{}_j \neq \delta^i_j \; , \tag{76}$$

where the Einstein summation convention has been used. Important differences arise between some of the expressions found here and those of standard Riemannian geometry because

$$\partial_l g^i{}_j \neq 0 \; . \tag{77}$$

Tensor fields on the manifold are mappings into the real numbers from outer products of the tangent space, T, and its dual vector space T^*. Furthermore, if a displacement is made along a zero eigenvector of the metric g_{ij}, this mapping must not change. This means that there are two neccesary and sufficient conditions for a general expression like $C^{ij}{}_{klm}(x)$ to be a tensor field (henceforth called a tensor). These are:

- **condition 1.** a tensor is not changed by contracting any index with that of the projection, e.g.

$$C^{ij}{}_{klm} = g^n{}_k C^{st}{}_{nrq} \ ,$$

and

$$C^{ij}{}_{klm} = g^i{}_s C^{st}{}_{nrq} \ .$$

- **condition 2.** defining the object (which lies in the space $T \otimes T \otimes T^* \otimes T^* \otimes T^*$)

$$\mathbf{C} \equiv C^{ij}{}_{klm} g^{kr} g^{ls} g^{mt} \mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_r \otimes \mathbf{e}_s \otimes \mathbf{e}_t \ ,$$

its derivative must satisfy

$$(\delta^l{}_m - g^l{}_m)\partial_l \mathbf{C} = 0 \ .$$

The metric tensor obviously satisfies condition 1. Consistency demands that any example must also satify condition 2. If this is the case, the covariant derivative of the metric tensor is zero, as will be shown below.

In Riemannian geometry, the covariant derivative of a tensor is the coordinate derivative of that tensor projected into tangent space. The definition of the covariant derivative on tensors with only upper indices (that is tensors which act on outer products of tangent space) is

$$\begin{aligned} \nabla_l A^{p_1 \cdots p_M} &= g^{p_1 k_1} \cdots g^{p_M k_M} \sum_{K_1 \ldots K_M} e^{K_1}{}_{k_1} \cdots e^{K_M}{}_{k_M} \\ &\times \partial_l (e^{K_1}{}_{m_1} \cdots e^{K_M}{}_{m_M} A^{m_1 \cdots m_M}) \ . \end{aligned} \tag{78}$$

For the case of a singular metric, $A_{p_1 , \ldots p_M}$ is not a tensor unless condition 2,

$$\nabla_j A^{p_1 \cdots p_M} = g^l{}_j \nabla_l A^{p_1 \cdots p_M} \ ,$$

is satisfied. Carrying out the differentiations in (78) yields

$$\begin{aligned} \nabla_l A^{p_1 \cdots p_M} &= g^{p_1}{}_{m_1} \cdots g^{p_M}{}_{m_M} \partial_l A^{m_1 \cdots m_M} \\ &+ \Gamma^{p_1}_{l m_1} g^{p_2}{}_{m_2} \cdots g^{p_M}{}_{m_M} A^{m_1 \cdots m_M} + \cdots \\ &+ g^{p_1}{}_{m_1} \cdots g^{p_{M-1}}{}_{m_{M-1}} \Gamma^{p_M}_{l m_M} A^{m_1 \cdots m_M} \ , \end{aligned} \tag{79}$$

where

$$\Gamma^p_{lm} = g^{pk} [\sum_K e^K{}_k \partial_l e^K{}_m] \ . \tag{80}$$

By the usual reasoning, (80) implies Christoffel's formula

$$\Gamma^p_{lm} = \frac{1}{2} g^{pk} (\partial_l g_{km} + \partial_m g_{kl} - \partial_k g_{lm}) \ .$$

The reader should notice that the covariant derivative above resembles the standard covariant derivative. However, because of (77), this is not true of the covariant derivative on a tensor with lower indices, $\nabla_l A_{k_1 \ldots k_M}$. This is

$$\begin{aligned}\nabla_l A_{k_1 \ldots k_M} &= \sum_{K_1 \ldots K_M} e^{K_1}{}_{k_1} \cdots e^{K_M}{}_{k_M} \partial_l (e^{K_1}{}_{m_1} \cdots e^{K_M}{}_{m_M} \\ &\times g^{m_1 r_1} \cdots g^{m_M r_M} A_{r_1 \ldots r_M}) \, .\end{aligned}$$

This may be written as

$$\begin{aligned}\nabla_l A_{k_1 \ldots k_M} &= g^{r_1}{}_{k_1} \cdots g^{r_M}{}_{k_M} \partial_l A_{r_1 \ldots r_M} \\ &- \Omega^{r_1}_{k_1 l}\, g^{r_2}{}_{k_2} \cdots g^{r_M}{}_{k_M} A_{r_1 \ldots r_M} - \cdots \\ &- g^{r_1}{}_{k_1} \cdots g^{r_{M-1}}{}_{k_{M-1}} \Omega^{r_M}_{k_M l}\, A_{r_1 \ldots r_M} \, ,\end{aligned} \tag{81}$$

where the affine connection now takes a different form:

$$\Omega^r_{kl} = -\sum_K e^K_k \partial_l (e^K{}_m g^{mr}) = -\partial_l g^r{}_k + \Gamma^r_{kl} \, , \tag{82}$$

which is not symmetric in the indices k and l. If $A_{r_1 \ldots r_M}$ is a tensor, by condition 2,

$$\nabla_j A_{k_1 \ldots k_M} = g^l{}_j \nabla_l A_{k_1 \ldots k_M} \, .$$

The covariant derivative of tensors with mixed upper and lower indices can be defined similarly. At any rate, indices of tensors can be lowered and raised by the metric tensor and its inverse, respectively.

There are more convenient expressions for covariant derivatives than those given above. Pushing the products of projections $g^{p_1}{}_{m_1} \cdots g^{p_M}{}_{m_M}$ past the partial derivative in (79) gives

$$\begin{aligned}\nabla_l A^{p_1 \cdots p_M} &= \partial_l A^{p_1 \cdots p_M} \\ &+ \Omega^{p_1}_{m_1 l}\, g^{p_2}{}_{m_2} \cdots g^{p_M}{}_{m_M} A^{m_1 \ldots m_M} + \cdots \\ &+ g^{p_1}{}_{m_1} \cdots g^{p_{M-1}}{}_{m_{M-1}} \Omega^{p_M}_{m_M l}\, A^{m_1 \ldots m_M} \, .\end{aligned} \tag{83}$$

Similiarly (81) can be rewritten as

$$\begin{aligned}\nabla_l A_{k_1 \ldots k_M} &= \partial_l A_{k_1 \ldots k_M} \\ &- \Gamma^{r_1}_{k_1 l}\, g^{r_2}{}_{k_2} \cdots g^{r_M}{}_{k_M} A_{r_1 \ldots r_M} - \cdots \\ &- g^{r_1}{}_{k_1} \cdots g^{r_{M-1}}{}_{k_{M-1}} \Gamma^{r_M}_{k_M l}\, A_{r_1 \ldots r_M} \, ,\end{aligned} \tag{84}$$

If the metric tensor is truly a tensor, meaning that it satisfies conditions 1 and 2 above, its covariant derivative vanishes. To prove this for singular metric tensors is a little more tedious than in the nonsingular case. Now

$$\begin{aligned}\nabla_l g_{k\,n} &= \partial_l g_{k\,n} - \Gamma^r_{k\,l} g_{r\,n} - \Gamma^s_{n\,l} g_{k\,s} \\ &= [\partial_l g_{k\,n} - \frac{1}{2} g^s_k \partial_l g_{s\,n} - \frac{1}{2} g^s_n \partial_l g_{s\,k}] \\ &- \frac{1}{2}[g^s_r \partial_k g_{s\,l} - g^s_k \partial_s g_{r\,l}] \\ &+ \frac{1}{2}[g^s_r \partial_s g_{k\,l} - g^s_k \partial_r g_{s\,l}] \, .\end{aligned} \tag{85}$$

Contracting this derivative with projection matrices must leave it unchanged:

$$\nabla_l g_{k\,n} = g^a_k g^b_n \nabla_l g_{a\,b} \ . \tag{86}$$

Applying (86) to each of the terms in brackets [ ] in (85) gives zero.

The first guess for the volume element is $\sqrt{det'\,g}$, where the prime means that the zero eigenvalue has been removed from the determinant. However, this cannot always be correct, because the integration $\int d^N x$ is over the entire coordinate space and not the d-dimensional manifold. In situations where the $N-d$ integration in the singular directions has a volume which is independent of the position on the manifold, this produces a (possibly infinite) constant factor, which can be divided out. For the example given in this appendix below and for the Yang-Mills theory, this turns out to be the case.

Consider a real-valued field $\phi(x)$. The derivative (which is the covariant derivative for a scalar), $\partial_k\phi$, and the projection of the derivative, $g^l{}_k\partial_l\phi$ must be the same, by condition 2 for the 0-tensor ϕ. The Laplacian is

$$\Delta\phi = -\nabla_j g^{jk}\partial_k\phi = -(\partial_j + \Omega^k_{j\,k})g^{j\,l}\partial_l\phi \ .$$

The Laplacian can be brought into a more convenient form using (73), (74) and (75). Using these expressions, the contraction of the connection $\Gamma^k_{j\,k}$ may be written as

$$\Gamma^k_{j\,k} = \sum_{J,\lambda_J\neq 0}(\frac{1}{2}\frac{\partial_j\lambda_J}{\lambda_J} + \sum_k S^J_k\partial_j S^J_k) = \frac{1}{\sqrt{det'\,g}}\partial_j\sqrt{det'\,g} + \frac{1}{2}\partial_j g^k_k \ .$$

Therefore

$$\Omega^k_{j\,k} = \frac{1}{\sqrt{det'\,g}}\partial_j\sqrt{det'\,g} + \frac{1}{2}\partial_j g^k_k - \partial_k g^k_j \ .$$

The Laplacian is consequently

$$\Delta\phi = -\frac{1}{\sqrt{det'\,g}}\partial_j(\sqrt{det'\,g}g^{j\,l}\partial_l\phi) + (\partial_k g^k_j - \frac{1}{2}\partial_j g^k_k)g^{j\,l}\partial_l\phi \ . \tag{87}$$

This is identical to the expression for the Laplacian in the nonsingular case, except for an additional term which depends on the derivatives of the projection matrix g^k_j.

The curvature is defined as the commutator of two covariant differentiations on a vector ξ^l. The definition does not need to be modified, because covariant differentiations of tensors are infinitesmal displacements in nonsingular directions (by condition 2 above). The Riemann curvature tensor is

$$R^k_{ijl}\xi^l = [\nabla_i, \nabla_j]\xi^k \ . \tag{88}$$

applying (83) and (84) to (88) gives

$$\begin{aligned} R^k_{ijl} &= \partial_i\Omega^k_{lj} - \partial_j\Omega^k_{li} + \Omega^k_{si}\Omega^s_{lj} - \Omega^k_{sj}\Omega^s_{li} \\ &= \partial_i\Gamma^k_{lj} - \partial_j\Gamma^k_{li} + \Omega^k_{si}\Omega^s_{lj} - \Omega^k_{sj}\Omega^s_{li} \ , \end{aligned}$$

where (82) was used in the last step. As usual, the the Ricci tensor is $R_{il} = R^k_{ikl}$, and the curvature scalar is $R = g^{il} R_{il}$.

As an application of these concepts, consider the two-dimensional unit sphere, with coordinates x^1, x^2, x^3, satisfying $-\infty < x^j < \infty$, excluding the point $\mathbf{x} = 0$. Then $\hat{\mathbf{x}} = \mathbf{x}/|\mathbf{x}|$ is a point on this sphere. The metric is the length of a chord connecting two points on the sphere. Thus if two three-vectors are $\mathbf{x}$ and $\mathbf{y}$, the distance between them, $s(\mathbf{x}, \mathbf{y})$, is given by

$$s(\mathbf{x}, \mathbf{y})^2 = (\frac{\mathbf{x}}{|\mathbf{x}|} - \frac{\mathbf{y}}{|\mathbf{y}|})^2 \ .$$

Thus

$$ds^2 = s(\mathbf{x} + d\mathbf{x}, \mathbf{x})^2 = g_{ij} dx^i dx^j$$

where the metric tensor is the three-by-three matrix

$$g_{ij} = \frac{1}{\mathbf{x} \cdot \mathbf{x}} (\delta_{ij} - \frac{x^i x^j}{\mathbf{x} \cdot \mathbf{x}}). \tag{89}$$

While lower indices are used on the (doubly-covariant) metric tensor, raised indices are used on the coordinates on the right-hand-side of (89). Since this equation is just an expression for the components of the metric tensor, this should cause no confusion. The reader can easily verify that the matrix g is singular. Three orthogonal eigenvectors are $\mathbf{x}$, any vector $\mathbf{y}$ which is perpendicular to $\mathbf{x}$ and $\mathbf{z} = \mathbf{x} \times \mathbf{y}$. The eigenvalues of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are 0, $(\mathbf{x} \cdot \mathbf{x})^{-1}$ and $(\mathbf{x} \cdot \mathbf{x})^{-1}$, respectively.

Notice that the manifold is *not* the two-dimensional real-projective space RP_2. In RP_2 the point $\mathbf{x} \in R^3$ is identified with $\lambda \mathbf{x}$ for any real $\lambda \neq 0$. The surface considered here is different in that λ is actually positive. The antipodes of the sphere are not identified.

It is straightforward to check that (89) gives the metric of the two-dimensional sphere after the substitution of polar coordinates r, θ and ϕ through

$$\mathbf{x} = r(\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta) \ .$$

The infinitesmal distance is then

$$ds^2 = d\theta^2 - \sin^2\theta d\phi^2 \ . \tag{90}$$

The geometric reason for the singularity of g is clear; the distance function does not depend on $\sqrt{\mathbf{x} \cdot \mathbf{x}} = r$. The meaning of this is that all the concentric spheres around the origin have been identified with one another.

At this stage, the reader might think that there is really no distinction between the singular metric formulation of the sphere and the standard embedding method of introducing a mapping

$$\theta, \phi \longrightarrow x^1, x^2, x^3 \ , \tag{91}$$

and defining the intrinsic geometric quantities in terms of the derivatives of this mapping. The new feature is that two-dimensional coordinate charts are never introduced.

In the usual approach, the embedding (91) is used to construct the normal bundle, which is a fiber bundle for the case of a sphere S_2 in a double covering of R^3 (as well as for the case of the manifold of gauge configurations). The singular metric technique makes it unneccesary to work with sections of this bundle; instead the points along each fiber are identified (much as physically-equivalent gauge fields are identified in this article).

The projection matrix is

$$g^i_j = \delta_{ij} - \frac{x^i x^j}{\mathbf{x}\cdot\mathbf{x}} \; , \tag{92}$$

which vanishes on the one-dimensional subspace spanned by $\mathbf{x}$ and gives unity on the space spanned by $\mathbf{y}$ and $\mathbf{z}$. This tensor plays the role of the identity. The reader who understands the mismatch of higher and lower indices in (89) should not have any objection in this regard to (92). The inverse metric tensor is

$$g^{ij} = \mathbf{x}\cdot\mathbf{x}\;\delta_{ij} - x^i x^j \; ,$$

and (76) is satisfied.

Naively, the volume element is $d^3x\sqrt{det'\,g}$. Such reasoning implies that the total volume is

$$V = \int d^3x\sqrt{det'\,g} = \int \frac{d^3x}{\mathbf{x}\cdot\mathbf{x}} = \infty \; ,$$

The problem is that singular directions are being integrated over, and the integration over these directions produces infinity. Now if the function to be integrated does not depend on $\mathbf{x}\cdot\mathbf{x}$, then there is a straightforward way to avoid doing this unwanted integration. The correct measure is obtained by dividing out the differential in the direction of the singular vector, $\mathbf{x}$:

$$dV = \frac{d^3x\sqrt{det'\,g}}{\mathbf{x}\cdot d\mathbf{x}} = 2\frac{d^3x\sqrt{det'\,g}}{d(\mathbf{x}\cdot\mathbf{x})} \; .$$

Formally, this is made somewhat easier by integrating this unwanted differential over a normalized function, so that the measure becomes, e.g.

$$dV = 2d^3x\sqrt{det'\,g}\;(\sqrt{\frac{\pi}{\alpha}}e^{-\alpha\mathbf{x}\cdot\mathbf{x}}) \; , \tag{93}$$

where α is a constant with units of length squared. Upon substituting polar coordinates, (93) is

$$dV = 2\sqrt{\frac{\pi}{\alpha}}\int_0^\infty dr\; e^{-\alpha r^2}\int_{-1}^{1} d\cos\theta\int_0^{2\pi} d\phi \; ,$$

which is clearly correct when integrating over functions which do not depend on r. Alternatively, if one is considering ratios of integrals on the sphere, the divergent factor from the unwanted integration (which is an overall constant) will cancel (though doing this explicitly requires a cut-off on the measure).

The affine connection (80) is

$$\Gamma^{j}_{k\,l} = \frac{2x^j x^k x^l}{(\mathbf{x}\cdot\mathbf{x})^2} - \frac{\delta_{j\,l}x^k + \delta_{j\,k}x^l}{\mathbf{x}\cdot\mathbf{x}} \ . \tag{94}$$

By constructing the vectors e^K_k, one can show that $e^J_l e^K_m g^{lm} = \delta^{JK}$, which is a constant. This makes explicit the fact that the covariant derivative of the metric tensor vanishes. This can also be seen by direct calculation:

$$\nabla_n g^j_q = g^r_q \nabla_n g^j_r = g^k_l g_{qs}\partial_n g^{ls} + g^{ks} g^r_q \partial_n g_{sr} = g^s_q(\partial_n g^p_s)g_{pq} \ . \tag{95}$$

Substituting (94) directly into (95) gives zero.

The Laplacian (87), with a little work, can be shown to agree with the expression of the Laplacian on a sphere in polar coordinates:

$$\Delta = -\frac{1}{\sin\theta}\partial_\theta \sin\theta\partial_\theta - \frac{1}{\sin^2\theta}\partial^2_\phi \ ,$$

which is left for the reader to verify.

Geodesics on the sphere are described by straight lines in R^3, namely $\mathbf{x}(t) = \mathbf{a}+\mathbf{b}t$, where $t \in (-\infty,\infty)$. Such a curve is mapped to a great half-circle by $\mathbf{x} \to \hat{\mathbf{x}}$. It is impossible to have a conjugate pair of points on this curve (these are antipodes on the sphere), since these are approached as $t \to \pm\infty$. However, by connecting several such curves together, a complete great circle can be made, and conjugate points can be reached.

Acknowlegements

I thank E. Akhmedov, J. Ambjørn, M. N. Chernodub, P. H. Damgaard, M. Kudinov, Y. Makeenko, M. Polikarpov, G. K. Savvidy and especially J. Huntley and V. P. Nair for discussions. Pierre van Baal pointed out a serious error concerning Chern-Simons integrals in an earlier version of this paper. Some of this work was done while visiting ITEP, whose staff I am grateful to for a very pleasant stay in Moscow. Finally, I thank the staff of the Niels Bohr Institute for their hospitality and financial support.

References

[1] R. P. Feynman, *Phys. Rev.* **80** (1950) 440.

[2] L. Dolan and R. Jackiw, *Phys. Rev.* **D9** (1974) 2904; K. Symanzik, *Nucl. Phys.* **B190** (1981) 1; T. Barnes and G.I. Ghandour, *Phys. Rev.* **D22** (1980) 925;

[3] J. P. Greensite, *Nucl. Phys.* **B158** (1979) 469; I. Kogan and A. Kovner, *Phys. Rev.* **D52** (1995) 3719.

[4] R. P. Feynman, *Nucl. Phys.* **B188** (1981) 479.

[5] I. M. Singer, *Physica Scripta* **24** (1980) 817.

[6] M. F. Atiyah, N. J. Hitchin and I. M. Singer, *Proc. R. Soc. Lond.* **A.362** (1978) 425.

[7] P. K. Mitter and C. M. Viallet, *Commun. Math. Phys.* **79** (1981) 457; *Phys. Lett.* **85B** (1979) 246; P. K. Mitter, in Proc. of NATO Adv. Study Inst., Cargèse, Aug. 1979, Plenum Press, New York (1980); M. S. Narasimhan and T. R. Ramadas, *Commun. Math. Phys.* **67** (1979) 21; M. Asorey and P. K. Mitter, *Commun. Math. Phys.* **80** (1981) 43.

[8] O. Babelon and C. M. Viallet, *Commun. Math. Phys.* **81** (1981) 515; *Phys. Lett.* **103B** (1981) 45.

[9] D. Karabali and V. P. Nair, *Nucl. Phys.* **B464** (1996) 135.

[10] D. Karabali and V. P. Nair, *Phys. Lett.* **379B** (1996) 141.

[11] D. Z. Freedman, P. E. Haagensen, K. Johnson and J. I. Latorre, MIT preprint MIT-CTP-2238 (1993) [hep-th 9309045]; P. E. Haagensen, K. Johnson and C. S. Lam, MIT preprint MIT-CTP-2492 (1995) [hep-th 9511226].

[12] Y. A. Simonov, *Sov. J. Nucl. Phys.* **41** (1985) 835; *Yad. Fiz.* **41** (1985) 1601.

[13] I. Chavel, "Eigenvalues in Riemannian Geometry", Academic Press (1984); E. B. Davies, "Heat Kernels and Spectral Theory", Cambridge Tracts in Mathematics **92**, Cambridge University Press (1989); H. Donnelly and P. Li, *Mich. Math. Journ.* **29** (1982) 149.

[14] M. Kudinov and P. Orland, in preperation.

[15] D. Zwanziger, *Nucl. Phys.* **B209** (1982) 336; G. Dell'Antonio and D. Zwanziger, *Nucl. Phys.* **B326** (1989) 333; *Commun. Math. Phys.* **138** (1991) 291; D. Zwanziger, *Nucl. Phys.* **B412** (1994) 657; M. Schaden and D. Zwanziger, contribution to Workshop on Quantum Infrared Physics, Paris (1994) [hep-th 9410019]

[16] P. Koller and P. van Baal, *Ann. Phys.* (N.Y.) **174** (1987) 299; *Nucl. Phys.* **B302** (1988) 1; P. van Baal, *Phys. Lett.* **224B** (1989) 397; *Nucl. Phys.* (proc. suppl.) **B17** (1990) 581; *Nucl. Phys.* **B351** (1991) 183; P. van Baal and N. D. Hari-Das, *Nucl. Phys.* **B385** (1992) 185; P. van Baal and B. van den Heuvel, *Nucl. Phys.* **B417** (1994) 215.

[17] R. E. Cutkosky, *J. Math. Phys.* **25** (1984) 939; R. E. Cutkosky and K. Wang, *Phys. Rev.* **D37** (1988) 3024; R. E. Cutkosky, *Czech. J. Phys.* **40** (1990) 252.

[18] K. Fujikawa, *Nucl. Phys.* **B468** (1996) 355; Tokyo University preprint UT-732 [hep-th 9511124]

[19] J. Fuchs, M. G. Schmidt and C. Schweigert, *Nucl. Phys.* **B426** (1994) 107.

[20] R. Friedberg, T. D. Lee, Y. Pang and H. C. Ren, *Ann. Phys.* (N.Y.) **246** (1996) 381.

[21] G. 't Hooft, *Phys. Rev. Lett.* **37** (1976) 8; *Phys. Rev.* **D14** (1976) 3432; R. Jackiw and C. Rebbi, *Phys. Rev. Lett.* **37** (1976) 1721; C. Callen, R. Dashen and D. Gross, *Phys. Lett.* **63B** (1976) 334.

[22] V. Gribov, *Nucl. Phys.* **B139** (1978) 1; I. M. Singer, *Commun. Math. Phys.* **60** (1978) 7.

[23] G. F. Simmons, "Introduction to Topology and Modern Analysis", McGraw-Hill, New York (1963); B. Epstein, "Linear Functional Analysis", Saunders, Philadelphia, PA (1970); M. Reed and B. Simon, "Methods of Modern Mathematical Physics", Vol. 1, Academic Press, New York (1972).

[24] A. D. Aleksandrov and V. A. Zalgaller, "Intrinsic Geometry of Surfaces", AMS Translations of Mathematical Monographs, **15**, ed. J. M Danskin (1967).

[25] M. A. Semenov-Tyan-Shanskii and V. A. Franke, *J. Sov. Math.* **34** (1986) 1999.

[26] K. Pohlmeyer, *Commun. Math. Phys.* **46** (1976) 207; A. M. Polyakov, *Phys. Lett.* **72B** (1977) 224; V. Zakharov and A. Mikhailov, *J.E.T.P.* **27** (1978) 47; M. Lüscher and K. Pohlmeyer, *Nucl. Phys.* **B137** (1978) 46; E. Brezin, C. Itzykson, J. Zinn-Justin and J.-B. Zuber, *Phys. Lett.* **82B** (1979) 442.

[27] G. Parisi, in "Recent Advances in Field Theory and Statistical Mechanics", Les Houches lectures 1982, ed. J.-B. Zuber and R. Stora, North-Holland (1984).

[28] G. K. Savvidy, *Phys. Lett.* **71B** (1977) 133; S. G. Martinian and G. K. Savvidy, *Nucl. Phys.* **B134** (1978) 539.

[29] B. Simon, *Ann. Phys.* **146** (1983) 209.

[30] G. 't Hooft, *Nucl. Phys.* **B72** (1974) 461.

[31] E. Witten, *Nucl. Phys.* **B149** (1979) 285.

[32] G. 't Hooft, *Nucl. Phys.* **B138** (1978) 1.

[33] M. Dine, C. Litwin and L. McLerran, *Phys. Rev.* **D23** (1981) 451.